```
                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:          November 24, 2003, 21:47:55 ; Search time 935 Seconds
                                               (without alignments)
                                               10831.397 Million cell updates/sec

Title:           US-10-001-857-42
Perfect score:   3096
Sequence:        1 ttcctcacgaaactcccagg...................atttctgacaaaaaaaaa 3096

Scoring table:   IDENTITY NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        2172232 seqs, 1635554964 residues

Total number of hits satisfying chosen parameters:        4344464

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
                10: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
                11: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
                12: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
                13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
                14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
                15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
                16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                    /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
c  17  58.2   1.9   3163   12   US-10-017-161-1857        Sequence 1857, Ap
c  18  58     1.9   2951   12   US-10-027-632-31508       Sequence 31508, A
c  19  58     1.9   956    12   US-10-027-632-31508       Sequence 31508, A
c  20  56.2   1.8   5452   12   US-10-017-161-1481        Sequence 1481, Ap
c  21  56     1.8   1117   12   US-10-017-161-1403        Sequence 1403, Ap
c  22  54.6   1.8   12733  14   US-10-032-393-47          Sequence 47, Appl
c  23  54.4   1.8   12739  14   US-10-032-393-47          Sequence 47, Appl
c  24  53.2   1.7   2561   10   US-10-023-529-48          Sequence 48, Appl
c  25  53.2   1.7   2561   10   US-10-023-529-48          Sequence 48, Appl
c  26  53.2   1.7   2561   13   US-10-023-529-48          Sequence 48, Appl
c  27  52.8   1.7   3133   12   US-10-017-161-1483        Sequence 1483, Ap
c  28  52.8   1.7   6432   12   US-10-329-079-10          Sequence 10, Appl
c  29  52.8   1.7   37360  12   US-10-329-079-6           Sequence 6, Appl
c  30  51.4   1.7   440    12   US-10-063-685-52          Sequence 52, Appl
c  31  51.4   1.7   440    14   US-10-184-644-202         Sequence 202, Appl
c  32  51.4   1.7   440    14   US-10-184-644-202         Sequence 202, Appl
c  33  51     1.6   152231 13   US-10-095-407-16          Sequence 16, Appl
c  34  50.6   1.6   65     12   US-10-908-975-879         Sequence 879, Appl
c  35  50.6   1.6   922    12   US-10-027-632-4358        Sequence 1435, App
c  36  50.4   1.6   883    12   US-10-027-632-4358        Sequence 4358, Ap
c  37  50.4   1.6   883    13   US-10-027-632-4358        Sequence 4358, Ap
c  38  50.2   1.6   1065   10   US-09-804-682-33          Sequence 33, Appl
c  39  50     1.6   1064   10   US-09-804-682-33          Sequence 33, Appl
c  40  50     1.6   1614   10   US-09-804-682-29          Sequence 29, Appl
c  41  50     1.6   1614   10   US-09-976-740-45          Sequence 45, Appl
c  42  50     1.6   1614   10   US-10-023-529-45          Sequence 45, Appl
c  43  50     1.6   12425  10   US-10-023-523-45          Sequence 45, Appl
c  44  50     1.6   12425  10   US-10-976-740-50          Sequence 50, Appl
c  45  50     1.6   12425  13   US-10-023-529-50          Sequence 50, Appl
                                US-10-023-523-50          Sequence 50, Appl
```

ALIGNMENTS

```
RESULT 1
US-10-001-857-42
; Sequence 42, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001,857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 42
; LENGTH: 3096
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-857-42

  Query Match       100.0%;  Score 3096;  DB 13;  Length 3096;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 3096;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy          1 TTCCTCACGAAACTCCCAGGCCGCGCTTGTATAGGAAACATAAATCCGTTGTGTCAGCAGTA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTCCTCACGAAACTCCCAGGCCGCGCTTGTATAGGAAACATAAATCCGTTGTGTCAGCAGTA 60

Qy         61 GCACGCGTGTTGCTCTCGGAGCTTGGCGTGCTCGTTCGTTGTTGTTGTTCGCCAGGTCTACG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACGCGTGTTGCTCTCGGAGCTTGGCGTGCTCGTTCGTTGTTGTTGTTCGCCAGGTCTACG 120

Qy        121 CAAACCTCACGGGTTTCCTTCCGCCCTTCGGCGTCGCCCTTTCTAAGAAATTCCCAGAGGGCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAAACCTCACGGGTTTCCTTCCGCCCTTCGGCGTCGCCCTTTCTAAGAAATTCCCAGAGGGCA 180
```

```
Qy  181  GCGCAGACGGGGCGGGGCCTCTGAGACTCCCGGGCTTCCGGGAAGCCCACTA  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  GCGCAGACGGGGCGGGGCCTCTGAGACTCCCGGGCTTCCGGGAAGCCCACTA  240

Qy  241  CCCAGAGACTCCCACAGAGAGGGTGAAAAAAGATAACTTCCGGTCTCGGATCGTCTCTAATC  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  CCCAGGACTCCCACAGAGGGGTGAAAAAGATAACTTCCGGTCTCGGATCGTCTCTAATC  300

Qy  301  TCGCGGAGAAGAGAAGGCCGGCCCGCATCGGCGGCGACCGGAGCGGGGGGTG  360
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  301  TCGCGGAGAAGAGAAGGCCCGCATCGGCGCCGACCGGAGCGGGGGGTG  360

Qy  361  TGGCCCGGGGAGGGGGTAAGGGGAGGGAGAGGGGGGCGGGGGGGGCCTCCCGG  420
         ||||||||||||||||||||||||||||||||||||||||
Db  361  TGGCCCGGGGAGGGGGTAAGGGGGAGGGAGAGGGGGGCGGGGGGGGCCTCCCGG  420

Qy  421  GCAATAGGCATGGCGACGCCGGGGGTCGGGGGTCGAGAGGGGCGCCGGTCCCGG  480
         |||||||||||||||||||||||||||||||||||||||||||||||
Db  421  GCATAGGCATGCCGTCACGCGCCGGGTCGGGGGCTGAGAGGGGCGCCGGTCCCGG  480

Qy  481  GCCGAGCGCGCGTCGTTATTTCCGTGCGGTCCGGACAGTGGGTGGGTGACCACG  540
         ||||||||||||||||||||||||||||||||||||||||||||
Db  481  GCCGAGGCGCGGCCGTCGTTTATTTCCGTGGTCCGGACAGTGGGTGACCACG  540

Qy  541  GGAGAAGTAGGCGATAATGGGTTATGAAAAGCTTCTGTTAGGATGATGAGGGGAG  600
         |||||||||||||||||||||||||||||||||||||||||||
Db  541  GGAGAAGTAGGCGATAATGGGTTATGAAAAGCTTCTGTAGATTGATGATGAGGGGAG  600

Qy  601  CTCAGTATGCCAGAAAAAATGGGAGAAAAGCAATACAAATGGGGTTGGACATTACCCAAGAT  660
         ||||||||||||||||||||||||||||||||||||||||||||
Db  601  CTCAGTATGCCAGAAAAAATGGGAGAAAAGCAATACAAATGGGGTTGGACATTACCCAAGAT  660

Qy  661  TTTGAAGAAGCTTGTCGAGAATTAAAGTTGGGGAGAACTTCAGATGATAAGCTATTTGGT  720
         |||||||||||||||||||||||||||||||||||||||||||
Db  661  TTTGAAGAAGCTTGTCGAGAATTAAAAGTTGGGGAGAACTTCAGATGATAAGCTATTTGGT  720

Qy  721  CTTTTTGAAGGCCATGTCTGCTATTGAAATGATGAATGGAATCCCAAGATCACTAGTC  780
         |||||||||||||||||||||||||||||||||||||||
Db  721  CTTTTGAAGGCCATGTCTGCTATTGAAATGATGAATGGAATCCCAAGATCACTAGTC  780

Qy  781  GGAAACCAAGTTAATGAAAAAGTTCTCAAATTTTGAACAAGGCACTATT  840
         |||||||||||||||||||||||||||||||||||||
Db  781  GGAAACCAAGTTAATCGAAAAGTTCTCAAATTTTGAACAAGGCACTATT  840

Qy  841  AAAATTAAAAGATCTCACCTTGCCTTGACTGATAGGGATTATGATACATGTTTTTTGGT  900
         ||||||||||||||||||||||||||||||||||||
Db  841  AAAATTAAAGATCTCACCTTGCCTTGACTGATAGGGATTATGATACATGTTTTGTGT  900

Qy  901  TTGATAACGTGGTTAGAAGGCCATTCACTGGCACAGCAGTATTTACGTGCCTTTACATT  960
         ||||||||||||||||||||||||||||||||||||||
Db  901  TTGATAACGTGGTTAGAAGGCCATTCACTGGCACAGCAGTATTTACGTGCCTTTACATT  960

Qy  961  CATAATCCAGACTTTATAGAAGAATCCTGCTATGAAAGGCTTTGCTCTGGGAATCTTGAAA  1020
         |||||||||||||||||||||||||||||||||||||
Db  961  CATAATCCAGACTTTATAGAAGAATCCTGCTATGAAAGGCTTTGCTCTGGGAATCTTGAAA  1020

Qy  1021 ATCTGTGACATTGCAAGGGAAAAAGTAAATAAGCTGCTGTTTTTGAAGGAGGAGTTTT  1080
         ||||||||||||||||||||||||||||||||||||
Db  1021 ATCTGTGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAGATTTT  1080

Qy  1081 CAGTCAAATGACTTATATGGATTTTAAAAATGGCTAACAGTGTGACAGTGACACTTGAGTTACAGGC  1140
         |||||||||||||||||||||||||||||||||||||
Db  1081 CAGTCAAATGACTTATGGATTTTAAAATGGCTAACAGTGTGACAGATCTTTGAGTTACAGGC  1140

Qy  1141 ATGCTAAAAGATGTGGGAGGGATGACATGTCAAAGAGTACTCGAAGTCGACAA  1200
         |||||||||||||||||||||||||||||||||||
Db  1141 ATGCTAAAAGATGTGGGAGGGATGACATGTCAAAGAGTACTCGAAGTCGACAA  1200

Qy  1201 GGAGAAGAAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGTTTAGCAGTATTCAGC  1260
         |||||||||||||||||||||||||||||||||||
Db  1201 GGAGAAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGTTTAGCAGTATTCAGC  1260
```

```
Qy  1261 AGAGTGAAATTACTCGTGTTGTTACTGACAGTGCTTATAGCCTTTACTTGAAGAGACC  1320
         ||||||||||||||||||||||||||||||||||||||
Db  1261 AGAGTGAAATTACTCGTGTTGTTACTGACAGTGCTTATAGCCTTTACTTGAAGAGACC  1320

Qy  1321 AGTGCTGTTGCAGAAGCTCAAAAAATGATGGTTCAAGCAGCAGAATCTTCTTTTCTGCCATT  1380
         |||||||||||||||||||||||||||||||||||||
Db  1321 AGTGCTGTTGCAGAAGCTCAAAAAATGATGGTTCAAGCAGCAGAATCTTCTTTTCTGCCATT  1380

Qy  1381 CATAATTCATTGCATCATGGCCATCAGGGCCTCCAGAAGTGAAGATCATCCA  1440
         |||||||||||||||||||||||||||||||||||
Db  1381 CATAATTCATTGCATCATGGCCATCAGGGCCTCCAGAAGTGAAGATCATCCA  1440

Qy  1441 ATTATGATTGGGTTTTGAACCCTTGTGAACCAGAGGGCTATTTTGCAAGATTAATAGATAGA  1500
         ||||||||||||||||||||||||||||||||
Db  1441 ATTATGATTGGGTTTTGAACCCTTGTGAACCAGAGGGCTATTTTGCAAGATTAATAGATAGA  1500

Qy  1501 TATGCAAAAAATAATTAAAAGGGAAGAAAAATGGTGAACTTAAATAGATTGCA  1560
         ||||||||||||||||||||||||||||||
Db  1501 TATGCAAAAAATAATTAAAAGGGAAGAAAAATGGTGAACTTAAATAGATTGCA  1560

Qy  1561 ATAAAAAACTGTCTGTGAAGTTGTGAAATTTAAACAAAATTACATTGTATCCTGGA  1620
         |||||||||||||||||||||||||||||
Db  1561 ATAAAAACTGTCTGTGAAGTTGTGAATTTAAACAAAATTACATTGTATCCTGGA  1620

Qy  1621 TGTGAATTAGTGAACAGTCACCATGTTCTTTAGATATCTGTTACAACTTTC  1680
         ||||||||||||||||||||||||||||||
Db  1621 TGTGAATTAGTGAACAGTCACCATGTTCTTTCAAGATATCTGTTACAACCTTTC  1680

Qy  1681 CTGGTGGATAACAAAAAAGTTCTTTGGAAACTCATCTCATGTGAAGGATGCA  1740
         |||||||||||||||||||||||||||||||
Db  1681 CTGGTGGATAACAAAAAAGGTCTTTGGAAACTCATCTCATGTGAAGGATGCA  1740

Qy  1741 CTTCCGTCTTTGTCAGATCCTCCGAGTGCCCGAGTCTACCTATATAATAATC  1800
         ||||||||||||||||||||||||||||||||
Db  1741 CTTCGGTTTGTCAGATCCTCCGAGTGCCCGAGTCTACCTATATAATAATC  1800

Qy  1801 ACCAGGCTAAGACTGTATCGACAGGAGTTATGACATTTGTTACTCACTGTCTCTC  1860
         |||||||||||||||||||||||||||||||
Db  1801 ACCAGGCTAAGACTGTATCGACAGGAGTTATGACATTTGTTACTCACTGTAGTC  1860

Qy  1861 TTATTCAGATCCATGGACAGTAACAGGGCTCGACAGAGAGTAAGCTTCATTCTTG  1920
         |||||||||||||||||||||||||||||||
Db  1861 TTATTCAGATCCATGGACAGTAACACAGGCTCGACAGAGAGTAAGCTTCATATTCTTG  1920

Qy  1921 AGGAATTGCCACCTTGCAGGATGAGTTATGACATTTATTTAATAGGCAGAAGGT  1980
         |||||||||||||||||||||||||||||
Db  1921 AGGAATTGCCACCTTGCAGGATGAGTTTAATGACATTTATTTAATAGGCAGAAGGT  1980

Qy  1981 TGATGCAGCGCTTCACCACCATGTGTTGAAAACAGGACAAACCCAACATTTGGGCCTG  2040
         ||||||||||||||||||||||||||||||
Db  1981 TGATGCAGCGCTTCACCACCATGTGTTGAAAACAGGAACCCAACATTTGGGCCTG  2040

Qy  2041 TTTAGGTACCTGGGTCCTTTACCATAACCTTGGCATTATGATAACTACCTTCTAAGTGG  2100
         |||||||||||||||||||||||||||||
Db  2041 TTTAGGTACCTGGGTCCTTTACCATAACCTTGGCATTATGATAACTACCTTCTAAGTGG  2100

Qy  2101 CTTTGAATTGGAACTCTACAGTATGCACGACTACTATTGGTACTCTCTGA  2160
         ||||||||||||||||||||||||||
Db  2101 CTTTGAATTGGAACTCTACAGTATGCACGACTACTATTGGTATCTCCTGA  2160

Qy  2161 ATTCCTTTACCCATGTTTGATTGCAGTACATTGACGTCGTTGCGTATGGGCAGA  2220
         |||||||||||||||||||||||||||||
Db  2161 ATTCCTTTACCGCATGGTTGATTGCAGTACATTGACGTCCCATGGCCGA  2220

Qy  2221 GGAAAAGGATAATGGAAGAGCAGCAGAAAGGCCGTAGTGTAAAAAAAAAA  2280
         ||||||||||||||||||||||||
Db  2221 GGAAAGGATAATGGAAGAGCAGCAGAAAGGCCCTAGTGTAAAAAAAAA  2280

Qy  2281 GAAAGTTCGCCCCATTGAGCCGAGAATCACAATGAGCCAAGAGCATATCAGATGTGTGC  2340
         ||||||||||||||||||||||||||
Db  2281 GAAAGTTCGCCCCATTGAGCCGAGAGATCACAATGAGCCAAGAGCATATCAGAACATGTGTGC  2340

Qy  2341 TGGAATGTTTAAAACCATGGGTAGCATTTGACATGGACGGCAAAGTACGGTTAAACCGAAGTT  2400
```

RESULT 2
US-09-814-353-20725

Sequence 20725, Application US/09814353
Publication No. US20030165831A1
GENERAL INFORMATION:
APPLICANT: Lee, John
APPLICANT: Thompson, Pamela
APPLICANT: Lillie, James
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
FILE REFERENCE: MRI-006B
CURRENT APPLICATION NUMBER: US/09/814,353
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: US 60/191,031
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: US 60/207,124
PRIOR FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: US 60/211,940
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: US 60/216,820
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: US 60/220,661
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: US 60/257,672
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 22037
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO: 20725
LENGTH: 2610
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1, 2, 3, 4
OTHER INFORMATION: n = A,T,C or G
US-09-814-353-20725

Query Match 72.0%;　Score 2228.4;　DB 12;　Length 2610;
Best Local Similarity 97.9%;　Pred. No. 0;　Mismatches 21;　Indels 28;　Gaps 3;
Matches 2299; Conservative 0; Mismatches 0;

Qy   426 CGCATGCGTGCCACGGTGCGCCGGGTGCGGGGCGTGGGCGTGGGTGAGGAGGGAGGGGCCGA 485
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2610 CGCATGCGTGCCACGCGTGCGCCGGGTGCGGGGCGTGGGCGTGGGTGAGGAGGGAGGGGCCGA 2551

Qy   486 GGCGGCGTGCGTTATTTCCGTGCGTCCGACAGTGCGTGCGGCCACGGGAGA 545
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2550 GGCGGCGTGCGTTATTTCCGTGCGTCCGACAGTGCGTGCGGCCACGGGAGA 2491

Qy   546 AGTAGGCATAATGGTTATGAAAAGCTTCGTGTAGATGATGACGATTCAG 605
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2490 AGTAGGCATAATGGTTATGAAAAGCTTCTGTAGATGATGACGATTCAGGAGCTCAG 2431

Qy   606 TATGCCAGAAAAAATGGAGAAAAGCAATACAAAACTGGGTGGCATTACCCAAGATTTTGA 665
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2430 TATGCCAGAAAAAATGGAGAAAAGCAATACAAAACTGGGTGGCATTACCCAAGATTTTGA 2371

Qy   666 AGAAGCTTGTCGACAATTAAAAGTTGGGAGAACTACTTCATGATGATAAGCTATTTGGTCTTTT 725
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2370 AGAAGCTTGTCGACAATTAAAAGTTGGGAGAACTACTTCATGATGATAAGCTATTTGGTCTTTT 2311

Qy   726 TGAAGCCATGTCTGCTATTGAAATGATGAGCCCAAGATGGATGCTGCCATGATTGGAAA 785
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2310 TGAAGCCATGTCTGCTATTGAAATGATGAGCCCAAGATGGATGCTGCCATGATTGGAAA 2251

Qy   786 CCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAAAT 845
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2250 CCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAAAT 2191

Qy   846 TAAAGATCTCACCTTGCCTGAACTGATAGGGATTATGGGATACACATGTTTTGCTTGTGGATT 905
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2190 TAAAGATCTCACCTTGCCTGAACTGATAGGGATTATGGGATACACATGTTTTGCTTGTGGAT 2131

Qy   906 AACGTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTTACGTGCCTTTACATTCATAA 965
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2130 AACGTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTTACGTGCCTTTACATTCATAA 2071

Qy   966 TCCAGACTTTTAGAGAGATCCCTGCTATGAAAGGCTTTGCTCCTGGGAATCTTGAAAATCTG 1025
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2070 TCCAGACTTTTATAGAGAGATCCCTGCTATGAAAGGCTTTGCTCCTGGGAATCTTGAAAATCTG 2011

Qy 1025 TGACATTGCAAGGGAGAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGAGATTTCAGTC 1085
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2010 TGACATTGCAAGGGAGAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGAGATTTCAGTC 1951

Qy 1086 AATGACTTATGCAGTTAAAAATGGCTAACAGTGTCAAAGACTTCAGGCATGCT 1145
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1950 AATGACTTATGGATTTAAAAATGGCTAACAGTGTGACAGAGTCTTCGAGTTACAGGCATGCT 1891

Qy 1146 AAAAGATGTTGGAGGATGACAGCAAAGAGAGAGTAAAGAGTACTCGAAGTCGACAAGGAGA 1205
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1890 AAAAGATGTTGGAGGATGACAGCAAAGAGAGAGTAAAGAGTACTCGGAAGTCGACAAGGAGA 1831

Qy 2341 TGGAATGTTTAAAACCATGGTAGCATTGACATTGACGGCACGGCAAAGTAACGTAAACCGAAGTT 2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2401 TGAGCTTGATAGTGAACAAGTTCGGTTATGAACACAGTTGCTCCATTCAACAGTGTGAT 2460

Qy 2401 TGAGCTTGATAGTGAACAAGTTCGGTTATGAACACAGTTGCTCCATTCAACAGTGTGAT 2460
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2461 GACCCGCGCGCCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAATAATAG 2520

Qy 2461 GACCCGCGCGCCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAATAATAG 2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2521 CCCTCCTCCTCAGTCTGTATGTGGCAGCTCAAGGAAAATGTCTCAATAATAAAATATAG 2580

Qy 2521 CCCTCCTCCTCAGTCTGTATGTGGCAGCTCAAGGAAAATGTCTCAATAATAAAATATAG 2580
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2581 AATGATATTGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTAAAAGGTTGC 2640

Qy 2581 AATGATATTGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTAAAAGGTTGC 2640
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2641 CAAACCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGCACACAAAAAGGAATCTAAAGT 2700

Qy 2641 CAAACCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGCACACAAAAAGGAATCTAAAGT 2700
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2701 TCCTCCTGAATTTGATTTCTCTGCTCATAAATATTTTCCTGTTGTGAAACTTGTTTTGAGA 2760

Qy 2701 TCCTCCTGAATTTGATTTCTCTGCTCATAAATATTTTCCTGTTGTGAAACTTGTTTTGAGA 2760
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2761 GAGACTGGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTCAGACCCAACTCTTAGAGGG 2820

Qy 2761 GAGACTGGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTCAGACCCAACTCTTAGAGGG 2820
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2821 CACATCACCAGCCTCCACATCACGGGGAGTGAGAGATGGATTTTCTTGGGTAACAACTCATTA 2880

Qy 2821 CACATCACCAGCCTCCACATCACGGGGAGTGAGAGATGGATTTTCTTGGGTAACAACTCATTA 2880
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2881 TAAGGAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAAACTGCTGTTTTGTTTAAAG 2940

Qy 2881 TAAGGAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAAACTGCTGTTTTGTTTAAAG 2940
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2941 TGGTTTATTATTGTTCCATGGAAGAAAACTGGTCTTATTGACATGAATGAAAACGTTATAT 3000

Qy 2941 TGGTTTATTATTGTTCCATGGAAGAAAACTGGTCTTATTGACATGAATGAAACGTTATAT 3000
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3001 GGTTTATTACAGATTTAATCACAAAATCATTTTTTATGGAATGAAATGAAAATAGTGT 3060

Qy 3001 GGTTTATTACAGATTTAATCACAAAATCATTTTTTATGGAATGAAATGAAAATAGTGT 3060
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 3061 TTATAAAGGTTAATAAATTTCTTGACAAAAAAAAA 3096

Qy 3061 TTATAAAGGTTAATAAATTTCTTGACAAAAAAAAA 3096

```
Qy  1206 AGAAAGAGAGATCCAGAAGTTGAACTAGAACACCAACAATGTTAGCGAGTATTCAGCGAGAGT 1265
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1830 AGAAAGAGAGATCCAGAAGTTGAACTAGAACACCAACAATGTTAGCGAGTATTCAGCGAGAGT 1771

Qy  1266 GAAATTTACTCGTGTTGTTACTCAGCAGTGCTTATAGCCTTTACTCAAGAAAAGAACACAGTGC 1325
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1770 GAAATTTACTCGTGTGTTACTCAGCAGTGCTTATAGCCTTTACTCAAGAAAAGAACACAGTGC 1711

Qy  1326 TGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGCAGATCTTCTTTCTGCCATTCATAA 1385
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1710 TGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGCAGATCTTCTTTCTGCCATTCATAA 1651

Qy  1386 TTCATTGCATCATGGCATGGCATCCAGAGCCCAGAAATGATACTACAAAAGAGGATCATCCAAATTAT 1445
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1650 TTCATTGCATCATGGCATCCAGAGCCCAGAAATGATACTACAAAAGAGAATCATCCAAATTAT 1591

Qy  1446 GATGGGTTTTGAACCCCTTGTGAACCAGAGCTACTCCACTCACCTTCCCTCCCTCGATATGC 1505
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1590 GATGGGTTTTGAACCCCTTGTGAACCAGAGCTACTCCACTCACCTTCCCTCCCTCGATATGC 1531

Qy  1506 AAAAATAATTAAAATGGGAAGAAATGGTGAACTATTTGCAGATTAATAGATAGAATAAA 1565
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1530 AAAAATAATTAAAATGGGAAGAAATGGTGAACTATTTGCAGATTAATAGATAGAATAAA 1471

Qy  1566 AACTGTCTCGTGAGGTGTGAATTTAACAAATTAACATGTATCCTGGATTTTTTCTGTGA 1625
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1470 AACTGTCTGTGAGGTGTGAATTTAACAATTTACATTGTATCCTGGATTTTTTCTGTGA 1411

Qy  1626 ATTTAGTGAACAGTCACCATGTCGTGTTCTTCAAGATCTCTTTACAACCACTTTCCTCGT 1685
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1410 ATTTAGTGAACAGTCACCATGTGTTCTTCAAGATCTCTTTACAACCACTTTCCTCGT 1351

Qy  1686 GGATAACAAAAAAGGTCTTTGGAACTCATCTCATCCATCCATATAATAATCACCAG 1745
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1350 GGATAACAAAAAAGGTCTTTGGAACTCATCTCATCGAAGACATGGTGAAAGATCCACTTCG 1291

Qy  1746 GTCTTTTTGTCAGATCCTCCGAGTGCTTCCCCAGTGCTACCTATATAATAATCACCAG 1805
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1290 GTCTTTTGTCAG--TCCTCCGGTGCTTCCCCAGTGCTACCTATATAATAATCACCAG 1233

Qy  1806 GCTAAGGACTGTATCGACTCCTTGTTACTCACTGTGTTCGGCCAATTCGTAGTCTTATT 1865
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1232 GCTAAGGACTGTATCGACTCCTTGTTACTCACTGTGTTCGGCCAATTCGTAGTCTTATT 1173

Qy  1866 CAGATCCATGGACATAACAGGGGCTTCGACAGAGATAAGCTTGGTCATATTCTTGAGGAA 1925
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1172 CAGATCCATGGACATAACAGGGGCTCGACAGAGATAAGCTTGGTCATATTCTTGAGGAA 1113

Qy  1926 TTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGCAGAGGAAGGTTGATG 1985
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1112 TTTGCCACCTTGCAGGATG--------AGGCAGAGGAAGGTTGATG 1076

Qy  1986 CAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAAGGCAACATTTGGCCTGTTTAG 2045
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1075 CAGCGCTTCACACCATGCTGTTGAAACAGGAACCCCAAAGGCAACATTTGGCCTGGTTAG 1016

Qy  2046 GTACCTGGGTCCTTTACCATAACCTTCGGATTATTGATACAGTACCTTCTTAAGTGGGCTTTG 2105
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1015 GTACCTGGGTCCTTTACCATAACCTTCGGATTATTGATACAGTACCTTCTTAAGTGGGTTTG 956

Qy  2106 AATTGGAACTCTACAGTATGCACGAGTACTATTACTACATGGTATCTCTCTGAATTCC 2165
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   955 AATTGGAACTCTACAGTATGCACGAGTACTATTACATATATTGGTATCTCTCTGAATTCC 896

Qy  2166 TTTACGCATGGTTGATGTCACACATTGAGTCGTGTGCCGATGGCTCTCAAATGGCAGAGGAAA 2225
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   895 TTTACGCATGGTTGATGTCACACATTGAGTCGTGTGCCGATGGCTCTCAAATGGCAGAGGAAA 836

Qy  2226 GGATAATGGAAGAGCAGCAAAGGCCGTAGTAGTGAAAAACAAAGAAAAAAAGAAAAG 2285
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   835 GGATAATGGAAGAGCAGAAAGGCCGTAGTAGTGAAAAACAAAGAAAAAAAGAAAAG 776

Qy  2286 TTCGCCCATTGAGCCGAGAGATCACAATGAGCCAAGCATATCAGAAACATGTGTGCTGGAA 2345
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   775 TTCGCCCATTGAGCCGAGAGATCACAATGAGCCAAGCATATCAGAAACATGTGTGCTGGAA 716

Qy  2346 TGTTTAAAACCATGGTAGACATTTGACATTGGACAGTTGACACAGTGAAAACCGAAAGTTGAGC 2405
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   715 TGTTTAAAACCATGGTAGACATTTGACATTGGACAGGGCAAAAGTACGTAAAACCGAAAGTTGAGC 656

Qy  2406 TTGATAGTGAACAAAGTTCGGTATGAACACAGAGGTTTGCTCATTCAACACAGTGTCCATTCAACA 2465
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   655 TTGATAGTGAACAAAGTTCGGTATGAACACAGAGGTTTGCTCATTCAACACAGTGTGATGACCC 596

Qy  2466 CGCGCGCCAGTGCACTACTTACAGTTCAAGGGAAAATGTCTGACCTCAATAATAATATAGCCCTC 2525
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   595 CGCGCGCCAGTGCACTACTTACAGTTCAAGGGAAAATGTCTGACCTCAATAATAATATAGCCCTC 536

Qy  2526 CTCCCTCAGTCTCCTGAACTGTATGTGGCCAGCTACTGAAGGAAATGTCGTAGTGACTTCAAGAGCCAAAATGA 2585
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   535 CTCCCTCAGTCTCCTGAACTGTATGTGGCCAGCTACTGAAGGAAATGTCGTAGTGACTTCAAGAGCCAAAATGA 476

Qy  2586 TATTGGAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATCTAAAGTTCCTC 2645
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   475 TATTGGAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTTAAAGTTGCCAAAC 416

Qy  2646 CCAACTTTGTGGTTATTGGGCAGGAGGACACAAAAAAGGAAATCTAAAGTTCCTC 2705
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   415 CCAACTTTGTGGTTATGAAGTTATTGGGCAGGAGGACACAAAAAAGGAAATCTAAAGTTCCTC 356

Qy  2706 CTGAATTTGATTTCTCTGCTCTATAAAATTATTTTCCTGTTGTTGTTGAGAGAGAC 2765
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   355 CTGAATTTGATTTCTCGTCATA----ATATTCCTGTGTTGAAACTGTTGAGAGAGACTG 299

Qy  2766 TGGGGAGG 2773
         ||||||||
Db   298 GGAGGTGG 291
```

Query Match   18.1%; Score 560.8; DB 12; Length 745;
Best Local Similarity 94.8%; Pred. No. 2.2e-141;
Matches 621; Conservative 0; Mismatches 8; Indels 26; Gaps 3;

```
Qy   1399 GGCATCCAGGCCCAGGATAGATACTACAGAAAAGGAGGATCATCCAATTATGATGGGTTTTGAA 1458
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    739 GGCATCCAGGCCCAGAATGATACTACAGAAAAGGAGGATCATCCAATTATGATGGGTTTTGAA  680
Qy   1459 CCCCTTGTGAACCAGGAGGCTACTTCCACCTACTTCCCTGGATATGCAAAAATAATTAAA 1518
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    679 CCCCTTGTGAACCAGGAGGCTACTTCCACCTACTTCCCTGGATATGCAAAAATAATTAAA  620
Qy   1519 AGGGAGAGAAATGGTGAACTATTTTGCAAGATTAATAGATAGAATAAAAACTGTCTGTGAG 1578
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    619 AGGGAGAGAAATGGTGAACTATTTTGCAAGATTAATAGATAGAATAAAAANCTGTCTGTGAG  560
Qy   1579 GTTGTGAATTTAACAAAATTTACATTGTATCCTGGAA-TTTTTTCTGTGAATTTAGTGAACA 1637
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    559 GTTGTGAATTTAACAAAATTTACATTGTATCCTGGATTTTTTTCTGTGAATTTAGTGAACA  500
Qy   1638 GTCACCATGTGTTCTTTCAAGATCTCTGTTACAAAACCACTTTCCTGGTGGATAACAAAAA 1697
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    499 GTCACCAGTGTGTTCTTTCAAGATCTCTGTTACAAAACCACTTTCCTGGTGGATAACAAAAA  440
Qy   1698 GGTGCTTTGGAACTCATCATGCTGAAAAGATGCACACTTCGGTCTCTTTTGTCAG 1757
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    439 GGTGCTTTGGAACTCATCATCAAGAACATAGGTGAAAAGATGCACTTCGGTCTCTTTTGTCAG  380
Qy   1758 ATCCTCCGAAGTGCTTTCCCCCAAGTGCTACCTTATATATAATAATCACCAGGCTAAGGACTGT 1817
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    379 TCCTCCG--GTGCTTTCCCCCAAGTGCTACCTTATATATAATAATCACCAGGCTAAGGACTGT  322
Qy   1818 ATCGACTCCTTTGTTACTCACTGTGTTCGGCCCATTCTTGTAGTCTTATTCAGATCCATGGA 1877
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 ATCGACTCCTTTGTTACTCACTGTGTTCGGCCCATTCTTGTAGTCTTATTCAGATCCATGGA  262
Qy   1878 CATAACAGGGCCTCGACAGAGAAGCTTGGTCATATATTCTTGAGGAATTTGGCCACCTTG 1937
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 CATAACAGGGCCTCGACAGAGAAGCTTGGTCATATTCTTGAGGAATTTGGCCACCTTG  202
Qy   1938 CAGGATGAGAGTTTATGACATTTTATTTAATTAGGCAGAGAGAGGTTGATGCAGCCGCTTCACA 1997
          |||||||||||||||||||||||||||||||||        ||||||||||||||||||||
Db    201 CAGGATG--------------------------AGGCAGAGAAGGTTGATGCAGCCGCTTCACA  165
Qy   1998 CCATGCTGTTGAAAACAGGAAACCCAAAGGGCAACACATTTTGGGCCTGTTTTAGGTACCTG 2052
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    164 CCATGCTGTTGAAAACAGGAAACCCAAAGGGCAACACATTTTGGGCCTGTTTTAGGTACCTG  110
```

RESULT 4
US-09-814-353-16155

- ; Sequence 16155, Application US/09814353
- ; Publication No. US20030165831A1
- ; GENERAL INFORMATION:
- ; APPLICANT: Lee, John
- ; APPLICANT: Thompson, Pamela
- ; APPLICANT: Lillie, James
- ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
- ; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
- ; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
- ; FILE REFERENCE: MRI-006B
- ; CURRENT APPLICATION NUMBER: US/09/814,353
- ; CURRENT FILING DATE: 2001-03-21
- ; PRIOR APPLICATION NUMBER: US 60/191,031
- ; PRIOR FILING DATE: 2000-03-21
- ; PRIOR APPLICATION NUMBER: US 60/207,124
- ; PRIOR FILING DATE: 2000-05-25
- ; PRIOR APPLICATION NUMBER: US 60/211,940
- ; PRIOR FILING DATE: 2000-06-15
- ; PRIOR APPLICATION NUMBER: US 60/216,820
- ; PRIOR FILING DATE: 2000-07-07
- ; PRIOR APPLICATION NUMBER: US 60/220,661

- ; PRIOR FILING DATE: 2000-07-25
- ; PRIOR APPLICATION NUMBER: US 60/257,672
- ; PRIOR FILING DATE: 2000-12-21
- ; NUMBER OF SEQ ID NOS: 22037
- ; SOFTWARE: FastSEQ for Windows Version 4.0
- ; SEQ ID NO 16155
- ; LENGTH: 610
- ; TYPE: DNA
- ; ORGANISM: Homo sapiens
US-09-814-353-16155

Query Match   16.5%; Score 509.4; DB 12; Length 610;
Best Local Similarity 99.8%; Pred. No. 1.8e-127;
Matches 510; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy   1184 GTACTCGAAGTCGACAAGGAGGAGAAAGAAGAGAAATCCAGAAGTTGAACTAGAACACAACAAT 1243
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    100 GTACTCGAAGTCGACAAGGAGGAGAAAGAAGAGAAATCCAGAAGTTGAACTAGAACACAACAAT  159
Qy   1244 GTTTAGCAGTATTCAGCAGAGTGAAATTTACTCCTGTTGTACGTGCTTATAGCCT 1303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    160 GTTTAGCAGTATTCAGCAGAGTGAAATTTACTCCTGTTGTACGTGCTTATAGCCT  219
Qy   1304 TTACTAAGAAGAGAGACCAAGGCTAGTCGTGGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAG 1363
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    220 TTACTAAGAAGAGAGACCAAGTGCTGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAG  279
Qy   1364 ATCTTTCTTTCTGCCATTCATAATTCATTGCATCATCAGCCGCCAGAATGATGATACTA 1423
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    280 ATCTTTCTTTCTGCCATTCATAATTCATTGCATCATCAGCCGCCAGAATGATGATACTA  339
Qy   1424 CAAAAGGAGAGCATCCAAATTTATGATGGGGTTTTGAAACCCCTTGTGAACCAGGAGGCTACTTC 1483
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    340 CAAAAGGAGAGCATCCAAATTTATGATGGGGTTTTGAAACCCCTTGTGAACCAGGAGGCTACTTC  399
Qy   1484 CACCTACTTCCCTCGATATGCAAAATAATAATTAATTAAAGAGGAAGAAAATGGTGAACTATTTTG 1543
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    400 CACCTACTTCCCTCGATATGCAAAATAATAATTAATTAAAGAGGAAGAAAATGGTGAACTATTTTG  459
Qy   1544 CAAGATTAATAGATAGAATAAAAACTGTCTGTGAGGTTGTGAATTTAACAAAATTTACATT 1603
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    460 CAAGATTAATAGATAGAATAAAAACTGTCTGTGAGGTTGTGAATTTAACAAAATTTACATT  519
Qy   1604 GTATCCTGGATTTTTTCTGTGAATTTAGTGAACAGTCACCACGTGTGTTCTTTCTTTCAAGATCTC 1663
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    520 GTATCCTGGATTTTTTCTGTGAATTTAGTGAACAGTCCCAGTGTGTTCTTTCTTTCAAGATCTC  579
Qy   1664 TGTTACAAAACCACTTTCCTGGTGGATAACAA 1694
          ||||||||||||||||||||||||||||||||
Db    580 TGTTACAAAACCACTTTCCTGGTGGATAACAA  610
```

RESULT 5
US-09-814-353-3218/c

- ; Sequence 3218, Application US/09814353
- ; Publication No. US20030165831A1
- ; GENERAL INFORMATION:
- ; APPLICANT: Lee, John
- ; APPLICANT: Thompson, Pamela
- ; APPLICANT: Lillie, James
- ; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
- ; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
- ; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
- ; FILE REFERENCE: MRI-006B
- ; CURRENT APPLICATION NUMBER: US/09/814,353
- ; CURRENT FILING DATE: 2001-03-21
- ; PRIOR APPLICATION NUMBER: US 60/191,031
- ; PRIOR FILING DATE: 2000-03-21
- ; PRIOR APPLICATION NUMBER: US 60/207,124
- ; PRIOR FILING DATE: 2000-05-25
- ; PRIOR APPLICATION NUMBER: US 60/211,940
- ; PRIOR FILING DATE: 2000-06-15
- ; PRIOR APPLICATION NUMBER: US 60/216,820

RESULT 6
US-09-814-353-9543/c
Sequence 9543, Application US/09814353
Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9543
; LENGTH: 602
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 561, 590
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-9543

Query Match   16.0%;   Score 496.2;   DB 12;   Length 602;
Best Local Similarity   93.9%;   Pred. No. 6.8e-124;
Matches 568; Conservative 0; Mismatches 10; Indels 27; Gaps 4;

```
Qy  1450 GGTTTTGAACCCCTTGTGAACACCAGAGCTACTTCCACCTACCTTCCCTCGATATGC-AAA 1508
         ||||||| ||| ||||||||||||||||| ||||||||||||||| |||||||||||  |||
Db   602 GGTTTTGAACCCNTTGTGAACCAGAGCGTACTTCCACCTACNTTCCCTCGATATGCAAAA  543

Qy  1509 AATAATTAAAAGGGAGAAAATGGTGAACTATTTTGCAAGATTAATAGATAGAATAAAAAC 1568
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   542 AATAATTAAAAGGGAGAAAATGGTGAACTATTTTGCAAGATTAATAGATAGAATAAAAACC 483

Qy  1569 TGTCTGTGAGGTTGTGAATTTAACAAATTTACAATTGTATCCTGGA-TTTTTTCTGTGAAT 1627
         |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   482 TGTCTGTGAGGTTGTGAATTTAACAAATTTACAATTGTATCCTGGATTTTTTCTGTGAAT  423

Qy  1628 TTAGTGAACAGTCACCATGTGTTCTTCAAGATCTCGTTACAAAACCACTTTCCTGGTGG 1687
         ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   422 TTAGTGAACAGTCACCACGTGTTCTTCAAGATCTCGTTACAAAACCACTTTCCTGGTGG  363

Qy  1688 ATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAAGATGCACTTCGGT 1747
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362 ATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAAGATGCACTTCGGT 303

Qy  1748 CTTTTGTCAGATCCTCCGAGTGCTTCCCCCAAGTGCTACTCCTATATAATAATCACCAGGC 1807
         |||||||||||   ||||||| |||| |||||||||||| ||||||||||||||||||||||
Db   302 CTTTTGTCAG--TCCTCCGGGTGCTTTCCCCAAGTGCTACCTATATAATAATCACCAGGC 245

Qy  1808 TAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTCGGCCATTCTGTAGTCTTATTCA 1867
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   244 TAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTCGGCCATTCTGTAGTCTTATTCA 185

Qy  1868 GATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAATT 1927
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   184 GATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAATT 125

Qy  1928 TGCCACCTTGCAGGATGAGTTTATGACATTTATTTAATAGGCAGAAGGTTGAGTGCA 1987
         ||||||||||||||||||              |||||||||||||||||||||||||
Db   124 TGCCACCTTGCAGGATGA-----------AGGCAGAAGGTTGAGTGCA 88

Qy  1988 GCGCTTCACACCATGCTGTTGAAACAGGACCCCAAAAGGCAACACATTTGGGCCTGTTTTAGGT 2047
         |||||||||||||||||||||
Db    87 GCGCTTCACACCATGCTGTTGAAACAGGACCCCAAAAGGCAACACATTTGGGCCTGTTTTAGGT 28

Qy  2048 ACCTG 2052
         |||||
Db    27 ACCTG 23
```

```
Db      87 GCGCTTCACACCATGCTCGTTGAAACAGGAACCCAAGGCAACACATTTGGCCTGTTTAGGT 28

Qy    2048 ACCTG 2052
           |||||
Db      27 ACCTG 23
```

RESULT 7
US-09-918-995-20914
;   Sequence 20914, Application US/0918995
;   Publication No. US20030073623A1
;   GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;   TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;   FILE REFERENCE: 20411-756
;   CURRENT APPLICATION NUMBER: US/09/918,995
;   CURRENT FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: US/09/235,076
;   PRIOR FILING DATE: 1999-01-20
;   NUMBER OF SEQ ID NOS: 38054
;   SOFTWARE: FastSEQ for Windows Version 3.0
US-09-918-995-20914
;   SEQ ID NO 20914
;   LENGTH: 482
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(482)
;   OTHER INFORMATION: n = A,T,C or G

Query Match   14.6%; Score 451.2; DB 11; Length 482;
Best Local Similarity   99.3%; Pred. No. 9.9e-112;
Matches   453; Conservative   0; Mismatches   3; Indels   0; Gaps   0;

```
Qy   491 CGTCGTTATTTTCCGTGTGTTCCGGACAGTGCTTGCCTGCTGACCACCGGGAGAAGTAG 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    26 CGAGCTTATTTCCGTGGTCGGACAGTGCCTGCCTGCTGACCACCGGGAGAAGTAG 85

Qy   551 GCATAATAGGTTATGAAAAGCTTCTCGTAGATGATGAGGAGGAGCTCAGTATGC 610
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    86 GCATAATGGTTATGAAAAGCTTCTCGTAGATGATGACGAATTCAGGATGC 145

Qy   611 CAGAAAAAATGGAGAAAAGCAATACAAAACTGGGGTGGCACATTACCCAAGATGTTTGAAGAAG 670
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   146 CAGAAAAAATGGAGAAAAGCAATACAAAACTGGGGTGGCACATTACCCAAGATTTTGAAGAAG 205

Qy   671 CTTGTCGGAGAATTAAAAGTTGGGGAGAACTACTCCATTCATGATAAGCTATTTGGTCTTTTGAAG 730
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   206 CTTGTCCAGAATTAAAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTTTGAAG 265

Qy   731 CCATGTCTGCTATTGAAAATGATGGATCCCAAGATGTTGGCATGAAAACCAAG 790
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   266 CCATGTCTGCTATTGAAAATGATGGATCCCAAGATGTTGGCATGAAAACCAAG 325

Qy   791 TTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCCAAGATGGAGATGGGAAAAATTAAAG 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   326 TTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGATGGCACTATTAAAAATTAAAG 385

Qy   851 ATCTCACCTTGCCTGAACTGATAGGGATTATAGGATTATACATGTTTTGCTGTTTGATAACGT 910
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   386 ATCTCACCTTGCCTGAACTGATAGGGATTATAGGATTATACATGTTTTGCTGTTTGATAACGT 445

Qy   911 GGTTAGAAGGCCATTCACTGGCACAGACAGCAGTATTTA 946
         ||||||||||||||||||||||||||||||||||||||
Db   446 GGTTAGAAGGCCATTCACTGGCACAGACAGCAGTATTTA 481
```

RESULT 8
US-09-814-353-3455
;   Sequence 3455, Application US/0914353
;   Publication No. US20030165831A1
;   Publication No. US20030165831A1

GENERAL INFORMATION:
;   APPLICANT: Lee, John
;   APPLICANT: Thompson, Pamela
;   APPLICANT: Lillie, James
;   GENERAL INFORMATION:
;   TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;   TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;   TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
;   FILE REFERENCE: MRI-006B
;   CURRENT APPLICATION NUMBER: US/09/814,353
;   CURRENT FILING DATE: 2001-03-21
;   PRIOR APPLICATION NUMBER: US 60/191,031
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: US 60/207,124
;   PRIOR FILING DATE: 2000-05-25
;   PRIOR APPLICATION NUMBER: US 60/211,940
;   PRIOR FILING DATE: 2000-06-15
;   PRIOR APPLICATION NUMBER: US 60/216,820
;   PRIOR FILING DATE: 2000-07-07
;   PRIOR APPLICATION NUMBER: US 60/220,661
;   PRIOR FILING DATE: 2000-07-25
;   PRIOR APPLICATION NUMBER: US 60/257,672
;   PRIOR FILING DATE: 2000-12-21
;   NUMBER OF SEQ ID NOS: 22037
;   SOFTWARE: FastSEQ for Windows Version 4.0
US-09-814-353-3455
;   SEQ ID NO 3455
;   LENGTH: 434
;   TYPE: DNA
;   ORGANISM: Homo sapiens

Query Match   14.0%; Score 434; DB 12; Length 434;
Best Local Similarity 100.0%; Pred. No. 4.3e-107;
Matches   434; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy  1186 ACTCGAAAGTCGACAAGGAGAGAGAAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGT 1245
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 ACTCGAAAGTCGACAAGGAGAGAGAAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGT 60

Qy  1246 TTAGCAGTATTCAGCAGAGAGTGAAAATTAACTCCGTGTGTTACTGACGAGTGCTTATAGCCTTT 1305
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TTAGCAGTATTCAGCAGAGAGTGAAAATTAACTCCGTGTGTTACTGACGAGTGCTTATAGCCTTT 120

Qy  1306 ACTAAGAAAGAGAGACCAGTGCTGCTGTTGCAGAAGCTCAAAAAATTGATGGTTCAAGCAGCGAGAT 1365
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACTAAGAAAGAGAGACCAGTGCTGTTGCAGAAGCTCAAAAAATTGATGGTTCAAGCAGCAGAGAT 180

Qy  1426 AAAGGAGATCATCCAATTATGAATGGGTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCA 1485
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AAAGGAGATCATCCAATTATGAATGGGTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCA 300

Qy  1486 CCTACCTTCCCTCGATATGCAAAAAATAATTAAAAGGGAGAAAATGGTGAACTATTTTTGCA 1545
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CCTACCTTCCCTCGATATGCAAAAATAATTAAAAGGGAGAAAATGGTGAACTATTTTTGCA 360

Qy  1546 AGATTAATTAGATAGAATAAAAAACTGTCTGTGGAGGTTGTTGAATTAACAAAATTTACATTGT 1605
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGATTAATTAGATAGAATAAAAAACTGTCTGTGGAGGTTGTTGAATTAACAAAATTTACATTGT 420

Qy  1606 ATCCTGGATTTTTT 1619
         |||||||||||||
Db   421 ATCCTGGATTTTTT 434
```

RESULT 9
US-09-814-353-9771
;   Sequence 9771, Application US/0914353
;   Publication No. US20030165831A1
;   GENERAL INFORMATION:
;   APPLICANT: Lee, John

Left column

```
;   APPLICANT: Thompson, Pamela
;   APPLICANT: Lillie, James
;   TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;   TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;   TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
;   FILE REFERENCE: MRI-006B
;   CURRENT APPLICATION NUMBER: US/09/814,353
;   CURRENT FILING DATE: 2001-03-21
;   PRIOR APPLICATION NUMBER: US 60/191,031
;   PRIOR FILING DATE: 2000-03-21
;   PRIOR APPLICATION NUMBER: US 60/207,124
;   PRIOR FILING DATE: 2000-05-25
;   PRIOR APPLICATION NUMBER: US 60/211,940
;   PRIOR FILING DATE: 2000-06-15
;   PRIOR APPLICATION NUMBER: US 60/216,820
;   PRIOR FILING DATE: 2000-07-07
;   PRIOR APPLICATION NUMBER: US 60/220,661
;   PRIOR FILING DATE: 2000-07-25
;   PRIOR APPLICATION NUMBER: US 60/257,672
;   PRIOR FILING DATE: 2000-12-21
;   NUMBER OF SEQ ID NOS: 22037
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 9771
;     LENGTH: 434
;     TYPE: DNA
;     ORGANISM: Homo sapiens
US-09-814-353-9771

  Query Match             14.0%;  Score 434;  DB 12;  Length 434;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-107;
  Matches  434;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   1186 ACTCGAAGTCGACAAGGAGAAGAAGAGATCCAGAGAAGTTGAACTAGAACACCAACAATGT 1245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ACTCGAAGTCGACAAGGAGAAGAAGAGATCCAGAGAAGTTGAACTAGAACACCAACAATGT 60

Qy   1246 TTAGCAGTATTCAGCAGAGAGTGAAAATTTACTCCGTGTGTTACTGACAGTGCTTATAGCCTTTT 1305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTAGCAGTATTCAGCAGAGAGTGAAAATTTACTCCGTGTGTTACTGACAGTGCTTATAGCCTTTT 120

Qy   1306 ACTAAGAAAAGAAGACCAGTGCTTGTGTTGCAGAAGCTCAAAAAAATTGATGGTTCAAGCAGCAGAT 1365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACTAAGAAAAGAAGACCAGTGCTTGTGTTGCAGAAGCTCAAAAAAATTGATGGTTCAAGCAGCAGAT 180

Qy   1366 CTTCTTCTTCTGCCATTCATTCATTGCCATCATGGCATCAGGCCCCAGAATGATAGTACTACA 1425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTTCTTCTTCTGCCATTCATTCATTGCCATCATGGCATCAGGCCCCAGAATGATAGTACTACA 240

Qy   1426 AAAGGAGATCATCCAATTATGATGGGTTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCA 1485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAAGGAGATCATCCAATTATGATGGGTTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCA 300

Qy   1486 CCTACTTCCCTCGATATGCAAAAAATAATTAAAAGGGAAGAAAATGGTGAACTATTTTGCA 1545
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCTACTTCCCTCGATATGCAAAAAATAATTAAAAGGGAAGAAAATGGTGAACTATTTTGCA 360

Qy   1546 AGATTAATAGATAGAATAAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTACATTGT 1605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGATTAATAGATAGAATAAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTACATTGT 420

Qy   1606 ATCCTGGATTTTTT 1619
          ||||||||||||||
Db    421 ATCCTGGATTTTTT 434
```

RESULT 10
US-09-983-965-210

```
;   Sequence 210, Application US/09983965
;   Patent No. US20020137160A1
;   GENERAL INFORMATION:
;   APPLICANT: Warren, Wesley C.
;   APPLICANT: Tao, Nengbing
;   APPLICANT: Byatt, John C.
```

Right column

```
;   APPLICANT: Mathialagan, Nagappan
;   TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;   TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;   FILE REFERENCE: 37-21(10297)C
;   CURRENT APPLICATION NUMBER: US/09/983,965
;   CURRENT FILING DATE: 2001-10-26
;   PRIOR APPLICATION NUMBER: US 09/465,231
;   PRIOR FILING DATE: 1999-12-15
;   PRIOR APPLICATION NUMBER: US 60/113,678
;   PRIOR FILING DATE: 1998-12-17
;   NUMBER OF SEQ ID NOS: 5912
;   SEQ ID NO 210
;     LENGTH: 398
;     TYPE: DNA
;     ORGANISM: Bos taurus
;     FEATURE:
;     OTHER INFORMATION: Clone ID: 25-BOVMS1-018-Q1-E1-G1
US-09-983-965-210

  Query Match             11.3%;  Score 350.8;  DB 10;  Length 398;
  Best Local Similarity   94.3%;  Pred. No. 1.7e-84;
  Matches 364;  Conservative   0;  Mismatches 22;  Indels   0;  Gaps   0;

Qy   2298 GCCGAGAGATCACAATTGAGCCAAGACATATCAGAACACATGTGTGCCTGGGAATGTTTAAAACCA 2357
          ||||||||||||||| ||||||||||  | |||||||||||||| ||| ||||||||||||||
Db     13 GCCGAGAGATCACAATGAGCCAAGCCAATATCAGAACACATGTGCTGCGAATGTTTAAAACCA 72

Qy   2358 TGGTAGCATTTGACATGGACGGCAAAGTACGTAAAACCCGAAGTTTGAGCTTGATAGTGAAC 2417
          |||||||||||||||||| ||||||| |||||||  |||||||||||||||||||| ||||
Db     73 TGGGAGCTTTTGACATGGACGGCAAAGTACGGAAGACCCGAAGTTTGAGCTTGATAGTGAAC 132

Qy   2418 AAGTTCGGTATGAAACACAGGTTTGCTCCATTCAACAGTGATGACACCCGCGCCAGTGC 2477
          ||||||||||||||| | ||||||||  |||||||||| || ||||||||||||||||
Db    133 AAGTTCGATATGAGCACAGATTTGCTCCATTCAACAGTGTAATGACACCCGCGCCAGTGC 192

Qy   2478 ACTACTTACAGTTCAAGGGAAAATGTCTGACCTCAATAAAATATATAGCCCTCAGTCTC 2537
          |||||| ||||||||||| |||||||||||||||||||||||| ||||||||||||||
Db    193 ACTATCTGCAGTTCAAGGGAGCTAGTTAATAATCTCGACCTCAATAAAATATAGCCCTCAGTCTC 252

Qy   2538 CTGAACTGTATGTGGGCAGCTAGTAAGCACTTCAACAGGCAAAAATGATATTGGAAAATA 2597
          || |||||| ||||||||||||||||||| ||||||||| ||||||||||||||||||
Db    253 CAGAACTGTATGTGGGCAGCTAGTAAGCACTTTCAGCAGGCAAAAATGATACTGGAAAATA 312

Qy   2598 TTCCTAACCCGACCATGAGGTTAATAGAATTTTAAAGGTTGCCAAACCCAACTTTGTGG 2657
          | |||||||||||| ||||||||| |||||||||||||| ||||||||||||||||||
Db    313 TCCCAAACCCAGACCATGAGGTCAATAGAATTTTAAAGGTTGCCAAACCCAACTTTGTGG 372

Qy   2658 TTATGAAGTTATTGGGCAGGAGGACAC 2683
          ||||||||||||||||||||||||||
Db    373 TTATGAAGTTATTGGGCAGGAGGACAC 398
```

RESULT 11
US-09-783-590-5693

```
;   Sequence 5693, Application US/09783590
;   Patent No. US20020108500A1
;   GENERAL INFORMATION:
;   APPLICANT: Dillon, Patrick J.
;   APPLICANT: Haseltine, William A.
;   APPLICANT: Li, Haodong
;   APPLICANT: Ruben, Steven M.
;   APPLICANT: Rosen, Craig A.
;   TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;   CURRENT APPLICATION NUMBER: US/09/783,590
;   CURRENT FILING DATE: 2000-02-15
;   FILE REFERENCE: PO-16.2C1
;   PRIOR APPLICATION NUMBER: 08/420,856
;   PRIOR FILING DATE: 1995-04-12
;   PRIOR APPLICATION NUMBER: 08/346,731
;   PRIOR FILING DATE: 1994-11-21
;   NUMBER OF SEQ ID NOS: 12485
;   SOFTWARE: Patentin Ver. 2.0
;   SEQ ID NO 5693
```

```
;  LENGTH: 501
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc feature
;  LOCATION: (226)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (289)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (315)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (353)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (381)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (383)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (390)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (403)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (416)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (425)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (428)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (440)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (449)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (455)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (466)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (479)
;  OTHER INFORMATION: n equals a,t,g, or c
;  NAME/KEY: misc feature
;  LOCATION: (487)
;  OTHER INFORMATION: n equals a,t,g, or c

US-09-783-590-5693

Query Match             8.7%;   Score 268.6;  DB 10;   Length 501;
Best Local Similarity   93.2%;   Pred. No. 4.5e-62;
Matches   313;  Conservative    0;  Mismatches   17;  Indels    6;  Gaps    3;

Qy    2223 AAAGGATAAATGGAGGAAGACAGCAGCAGAAAGGCCGTAGTAGTAGTAAAAAAACAAAGAAAAAAAGA 2282
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7 AAAGGATAAATGGAGGAAGACAGCAGCAGAAAGGCCCTAGTAGTAGTAAAAAAACAAAGAAAAAAAGA 66

Qy    2283 AAGTTCGCCCATTGAGCCGGAGAGATCACACATGAGCGAAGCCATTATCACGAACACATGTGTCTG 2342
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      67 AAGTTCGCCCATTGAGCCGCGAGAGATCACAATCAATCACGAACACATGTGTGCTG 126

Qy    2343 GAATGTTTAAAACCATGGTAGCATTTGACTACGTAAACCGAAGTTTG 2402
           ||||||||||||||||||||||||||||||||||||||||||||||
Db     127 GAATGTTTAAAACCATGGTTAGCATTTGACATGGACGGCAAAGTACGTAAACCGAAGTTTG 186


RESULT 12
US-10-027-632-323193/c
Sequence 323193, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 323193
;  LENGTH: 489
;  TYPE: DNA
;  ORGANISM: Human
US-10-027-632-323193

Query Match             5.9%;   Score 182;  DB 12;   Length 489;
Best Local Similarity   100.0%;   Pred. No. 1.6e-38;
Matches   182;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1431 AGATCATCCAATTATGATGGGTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACTAC 1490
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     401 AGATCATCCAATTATGATGGGGTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACTAC 342

Qy    1491 CTTCCCTCGATATGCAAAAATAATTAAAAAGGAGAAAATGGTGAACTACTATTTTGCAAGATT 1550
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     341 CTTCCCTCGATATGCAAAAATAATTAAAAAGGAGAAAATGGTGAACTACTATTTTGCAAGATT 282

Qy    1551 AATAGATAGAATAAAAAACTGTCTTGTGAGGTTGTAGAATTTAACAAATTTACATTGTATCCT 1610
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 AATAGATAGAATAAAAAACTGTCTTGTGAGGTTGTAGAATTTAACAAATTTACATTGTATCCT 222

Qy    1611 GG 1612
           ||
Db     221 GG 220


RESULT 13
US-10-027-632-323193/c
; Sequence 323193, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```
;   TITLE OF INVENTION: Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR APPLICATION FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSEQ for Windows Version 4.0
;
;   SEQ ID NO 323193
;   LENGTH: 489
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-323193

Query Match             5.9%;  Score 182;  DB 13;  Length 489;
Best Local Similarity   100.0%;  Pred. No. 1.6e-38;
Matches 182;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1431 AGATCATCCAATTATGATGAGGGTTTTGAACCCTTGTGTAACCAGAGGCTACTTCCACCTAC 1490
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 AGATCATCCAATTATGATGAGGGTTTTGAACCCTTGTGTAACCAGAGGCTACTTCCACCTAC  342

Qy      1491 CTTCCCTCGATATGCAAAAATAATTAAAAGGGAAGAGAAAATGGTGAACTATTTTGCAAGATT 1550
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 CTTCCCTCGATATGCAAAAATAATTAAAAGGGAAGAGAAAATGGTGAACTATTTTGCAAGATT  282

Qy      1551 AATAGAATAGAATAAAAACTGTCTCGTGTGAATTTAACAAATTTACATTGTATCCT 1610
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 AATAGAATAGAATAAAAACTGTCTCGTGTGAATTTAACAAATTTACATTGTATCCT  222

Qy      1611 GG 1612
             ||
Db       221 GG  220
```

```
Matches 141;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      1911 CATATATCTTGAGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTATTTAATAGG 1970
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       151 CATATTCTTGAGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTATTTAATAGG   92

Qy      1971 CAGAGAAGGTTGAATGCAGCGCCTTCACACCATGCTTGTTGAAACAGGAACCCCAAAGGCAAC 2030
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        91 CAGAGAAGGTTGAATGCGCGCCTTCACACCATGCTTGTTGAAACAGGAACCCCAAAGGCAAC   32

Qy      2031 ATTTGGCCTGTTTAGGTACCTG 2052
             ||||||||||||||||||||||
Db        31 ATTTGGCCTGTTTAGGTACCTG   10
```

RESULT 14
US-10-001-857-41/c
; Sequence 41, Application US/10001857
; Publication No. US20201835000A1
; GENERAL INFORMATION:
;   APPLICANT: Macina, Roberto
;   APPLICANT: Recipon, Herve
;   APPLICANT: Chen, Sei-Yu
;   APPLICANT: Sun, Yongming
;   APPLICANT: Liu, Chenghua
;   TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
;   FILE REFERENCE: DEX-0273
;   CURRENT APPLICATION NUMBER: US/10/001,857
;   CURRENT FILING DATE: 2001-11-20
;   PRIOR APPLICATION NUMBER: 60/252,054
;   PRIOR FILING DATE: 2000-11-20
;   NUMBER OF SEQ ID NOS: 208
;   SOFTWARE: Patentin version 3.1
;
;   SEQ ID NO 41
;   LENGTH: 151
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-001-857-41

Query Match   4.5%; Score 140.4; DB 13; Length 151;
Best Local Similarity   99.3%; Pred. No. 1.3e-27;

RESULT 15
US-10-001-857-111/c
; Sequence 111, Application US/10001857
; Publication No. US20201835000A1
; GENERAL INFORMATION:
;   APPLICANT: Macina, Roberto
;   APPLICANT: Recipon, Herve
;   APPLICANT: Chen, Sei-Yu
;   APPLICANT: Sun, Yongming
;   APPLICANT: Liu, Chenghua
;   TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P
;   FILE REFERENCE: DEX-0273
;   CURRENT APPLICATION NUMBER: US/10/001,857
;   CURRENT FILING DATE: 2001-11-20
;   PRIOR APPLICATION NUMBER: 60/252,054
;   PRIOR FILING DATE: 2000-11-20
;   NUMBER OF SEQ ID NOS: 208
;   SOFTWARE: Patentin version 3.1
;
;   SEQ ID NO 111
;   LENGTH: 815
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (206)..(305)
;   OTHER INFORMATION: a, c, g or t
US-10-001-857-111

Query Match   2.8%; Score 87.8; DB 13; Length 815;
Best Local Similarity   83.0%; Pred. No. 9.4e-13;
Matches 122; Conservative   0; Mismatches   2; Indels   23; Gaps   1;

```
Qy      1910 TCATATTCTTGAGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTATTTAATAG 1969
             ||||||||||||||||||||||||||||||||||||||||||||||         ||||
Db       205 TCATATTCTTGAGGAATTTGCCACCTTGCAGGATG---------------AG  169

Qy      1970 GCAGAGAAGGTTGATGCAGCGCGTTCACACCATGCTTGTTGAAACAGGAACCCCAAAGGCAA 2029
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       168 GCAGAGAAGGTTGATGCAGCGCTTCACACCATGCTTGTTGAAACAGGAACCCCAAAGGCAA  109

Qy      2030 CATTTGGCCTGTTTAGGTACCTGGGGTC 2056
             ||||||||||||||||||||||||||||
Db       108 CATTTGGCCTGTTTAGGTACCTGGGCC   82
```

Search completed: November 25, 2003, 03:08:51
Job time : 938 secs

OM nucleic - nucleic search, using sw model

Run on: November 24, 2003, 21:45:44

Title: US-10-001-857-42
Perfect score: 3096
Sequence: 1 ttcctcacgaaactcccagg..............atttcttgacaaaaaaaaa 3096

Scoring table: IDENTITY_NUC
                     Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq 19Jun03:*
```
 1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
 2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
 3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
 4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
 5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
 6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
 7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
 8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
 9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

; Search time 779 Seconds
(without alignments)
10728.447 Million cell updates/sec

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3096	100.0	3096	24	ABQ75303	Human lung specifi
c 2	2577.8	83.3	2668	23	ABV23940	Human prostate exp
c 3	2577.8	83.3	2668	23	ABV29823	Human prostate exp
c 4	2417.2	78.1	2488	24	AAL49929	Human molecule for
5	717.4	23.2	719	24	ABS51309	cDNA encoding huma
6	350.8	11.3	398	25	ABX50281	Bovine EST associa
c 7	343.8	11.1	799	24	ABQ51658	Oligonucleotide fo
c 8	343.8	11.1	799	24	ABQ51659	Oligonucleotide fo

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
 c  9  292.4   9.4   356   23  ABV15331   Human prostate exp
 c 10  288.8   9.3   439   23  ABV36125   Human prostate exp
 c 11  288.8   9.3   439   23  ABV45182   Human prostate exp
 c 12  274.2   8.9   799   24  ABQ51660   Oligonucleotide fo
 c 13  274.2   8.9   799   24  ABQ51661   Oligonucleotide fo
 c 14  230.2   7.4   303   23  ABN43747   Human spliced tran
    15  217     7.0   2515  23  ABV06162   Human prostate exp
    16  140.4   4.5   151   24  ABL07143   Human lung specifi
 c 17  90.6    2.9   600   24  ABQ75302   Oligonucleotide fo
 c 18  90.6    2.9   600   24  ABQ52497   Oligonucleotide fo
 c 19  87.8    2.8   4847  24  ABL05906   Primer used in the
 c 20  60      1.9   815   24  ABL07142   Human lung specifi
    21  60      1.9   60    24  ABQ75372   Human lung specifi
 c 22  58.6    1.9   600   24  ABS24496   Oligonucleotide fo
 c 23  58.6    1.9   600   24  ABQ52497   Oligonucleotide fo
 c 24  56.4    1.8   320   21  AAA38183   Primer used in the
 c 25  56.4    1.8   320   21  ABT10409   Human breast cance
    26  56      1.8   1286  24  ABI99656   Mouse ischaemic co
    27  55.4    1.8   365   23  ABV54856   Human prostate exp
 c 28  55.2    1.8   320   21  AAA38186   Primer used in the
    29  54.8    1.8   840   24  AAA38185   Primer used in the
 c 30  54.6    1.8   840   24  ABQ35494   Oligonucleotide fo
    31  54.6    1.8   840   24  ABQ35495   Oligonucleotide fo
    32  54.6    1.8   12733 24  ABZ20967   Animal test kit ma
    33  54.6    1.8   12733 24  ABK98631   Vector pPEPF14 con
    34  54.2    1.8   12739 24  ABK98592   Vector pPEPF1 cont
    35  54.2    1.8   318   21  AAA38184   Primer used in the
 c 36  54.2    1.8   4403765 22 AAI99683   Mycobacterium tube
    37  53.8    1.8   4411529 22 AAI99682   Mycobacterium tube
 c 38  53.4    1.7   2188   20  AAZ77506   Human ovarian tumo
    39  53.4    1.7   1337   20  AAZ17263   Human gene express
    40  53      1.7   2561   22  AAH26500   Rabbit low density
 c 41  52.8    1.7   434    22  AAI84343   Human polynucleoti
    42  52.8    1.7   344    22  AAI82007   Human polynucleoti
    43  52.2    1.7   1327   24  ABQ68452   Listeria monocytog
    44  51.6    1.7   1000   21  AAA02484   Human colon cancer
 c 45  51.4    1.7   237    21  AAA81807   N. meningitidis pa
    45  51      1.6   712    24  ABQ40858   Oligonucleotide fo
```

ALIGNMENTS

RESULT 1
ABQ75303

ID   ABQ75303 standard; cDNA; 3096 BP.
XX
AC   ABQ75303;
XX
DT   05-NOV-2002 (first entry)
XX
DE   Human lung specific nucleic acid sequence SEQ ID NO:42.
XX
KW   Human; lung; lung specific nucleic acid; LSNA; lung specific protein;
KW   LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;
KW   squamous cell carcinoma; gene; chromosome 9; ss.
XX
OS   Homo sapiens.
XX
PN   WO200264788-A2.
XX
PD   22-AUG-2002.
XX
PF   20-NOV-2001; 2001WO-US045080.
XX
PR   20-NOV-2000; 2000US-252054P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Macina RA, Recipon H, Chen S, Sun Y, Liu C;
XX
DR   WPI; 2002-657601/70.
XX
PT   New lung specific nucleic acid useful in gene therapy or as vaccines

```
PT   for treating lung cancer (e.g. squamous cell carcinoma) or
PT   non-cancerous lung diseases, as well as for diagnosing, monitoring or
PT   staging these diseases -
XX
PS   Claim 1; Page 168-169; 282pp; English.
XX
CC
CC   The present invention describes an isolated lung specific nucleic acid
CC   (LSNA) comprising a sequence that: (a) encodes any of the 93 amino acid
CC   sequences comprising 17 - 733 amino acids, given in ABPS2873 to ABPS2965;
CC   (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),
CC   given in ABQ75262 to ABQ75376; (c) selectively hybridises to (a) or (b);
CC   or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific
CC   protein (LSP) sequences have cytostatic activity and can be used in gene
CC   therapy and vaccines. LSNAs and LSPs are useful for diagnosing and
CC   monitoring the presence and metastases of lung cancer in a patient. An
CC   antibody that specifically binds to an LSP can be used for determining
CC   the presence of an LSP in a sample, as well as for treating a patient
CC   with lung cancer, particularly by inducing an immune response against
CC   the lung cancer cell expressing the LSNAs or LSPs. In particular, these
CC   LSNAs and LSPs are useful for identifying, diagnosing, monitoring,
CC   staging, imaging and treating lung cancer (e.g. squamous cell carcinoma)
CC   and non-cancerous disease states in lung.
XX
SQ   Sequence 3096 BP; 914 A; 606 C; 749 G; 827 T; 0 other;

Query Match             100.0%;  Score 3096;  DB 24;  Length 3096;
Best Local Similarity   100.0%;
Matches 3096;  Conservative     0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy          1 TTCCTCACGAAACTCCCAGCGCGCTGTATAGGAAACAATAAATCCGTTGTCAGGCAGCAGTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTCCTCACGAAACTCCCAGCGCGCTGTATAGGAAACAATAAATCCGTTGTCAGGCAGCAGTA 60

Qy         61 GCACGCTGTTGCTCTCGGAGCTTGCTGCTTCGTTCGTTCGTCCGAACCACTAAGGTCTACG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCACGCTGTTGCTCTCGGAGCTTGCTGCTTCGTTCGTTCGTCCGAACCACTAAGGTCTACG 120

Qy        121 CAAACCTTCACGGTTTCCTTCCGCGCCTTCGCGTCACCTTTCTAAGAAAATTCCCAGAGGCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CAAACCTTCACGGTTTCCTTCCGCGCCTTCGCGTCACCTTTCTAAGAAAATTCCCAGAGGCA 180

Qy        181 GCGGAGACGGGGGCGGGGGCGGGGGCGGGCTCTGAGACTCCGGGGAACGCGCCCACTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCGGAGACGGGGGCGGGGGCGGGGGCGGGCTCTGAGACTCCGGGGAACGCGCCCACTA 240

Qy        241 CCCAGGAGCTCCGACAGAGGGTGAAAAAAGATAACTTCCGGTCTCGCGATGCGTCTCTAATC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCCAGGAGCTCCGACAGAGGGTGAAAAAAGATAACTTCCGGTCTCGCGATGCGTCTCTAATC 300

Qy        301 TCGCGAGAGAGAAGGCCCGGCCCATCGGCCGGGGCCGAGGGGTGCGAGGGAGGGGGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCGCGAGAGAGAAGGCCCGGCCCATCGGCCGGGGCCGAGGGGTGCGAGGGAGGGGGTG 360

Qy        361 TGGCCCGGGGGAGGGGAGTGAGGGAGGACAGGGGGGGGGTCCCGGGAGTGAGGGAGG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGGCCCGGGGGAGGGGAGTGAGGGAGGACAGGGGGGGGGTCCCGGGAGTGAGGGAGG 420

Qy        421 GCATACGGCATGCGGTGCACGCTGACGCGCGGGTGCGCGGGGGCGGGGGGGCGGGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GCATACGGCATGCGGTGCACGCTGACGCGCGGGTGCGCGGGGGCGGGGGGGCGGGCG 480

Qy        481 GCCGAGGGCGGCGTCGTTATTTCCGGTGGTCCGGACCAGTGACGGGGTTGACCACG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCGAGGGCGGCGTCGTTATTTCCGGTGGTCCGGACCAGTGACGGGGTTGACCACG 540

Qy        541 GGAGAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGACGATTCAGGATGGGAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGAGAAGTAGGCATAATGGTTATGAAAGCTTCTGTAGATGACGATTCAGGATGGGAG 600

Qy        601 CTCAGTATGCCAGAAAAAAATGGAGAAAAAGCAATACAAACTGGGTGGACATTACCCAAGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTCAGTATGCCAGAAAAAAATGGAGAAAAAGCAATACAAACTGGGTGGACATTACCCAAGAT 660
```

```
Qy        661 TTTGAAGAAGCTTGTGTCCGAGAATTAAAGTTGGGAGAACTACTTCAGTGATAAAGCTATTTGGT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TTTGAAGAAGCTTGTGTCCGAGAATTAAAGTTGGGAGAACTACTTCAGTGATAAAGCTATTTGGT 720

Qy        721 CTTTTTGAAGCCATGTCTGCTATTGAAGATTGAAAATGATGGAAGATCCCAAGATGCTTGGCATGATT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTTTTTGAAGCCATGTCTGCTATTGAAGATTGAAAATGATGGAAGATCCCAAGATGCTTGGCATGATT 780

Qy        781 GGAAACCAAGTTAATCGAAAAGTTCTCAATTTTGAAGATGATGGCACTATTCCAAGATGGGAGCACAA 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGAAACCAAGTTAATCGAAAAGTTCTCAATTTTGAAGATGATGGCACTATTCCAAGATGGGAGCACAA 840

Qy        841 AAAATTAAAGATCTCACCTTGCCTGAACTGATCAGGATTATGGATACAATGTTTTGCTGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAAATTAAAGATCTCACCTTGCCTGAACTGATCAGGATTATGGATACAATGTTTTGCTGT 900

Qy        901 TTGATAACGTTGGTTAGAGGCCATTCACTGGCACAGCAGTATTTACGGTGCCTTTACATT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTGATAACGTTGGTTAGAGGCCATTCACTGGCACAGCAGTATTTACGGTGCCTTTACATT 960

Qy        961 CATAATCCAGACACTTTATAGGAAGATCCTGCTATGAAGAGGCTTTTGCTCTGGGAATCTGAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CATAATCCAGACACTTTATAGGAAGATCCTGCTATGAAGAGGCTTTTGCTCTGGGAATCTGAAA 1020

Qy       1021 ATCTGTGACATTGCAAGGGAAAAAGTAAAATAAAAAGCTGCTCTGTTTTTTGAAGAGGAGATTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ATCTGTGACATTGCAAGGGAAAAAGTAAAATAAAAAGCTGCTCTGTTTTTTGAAGAGGAGATTT 1080

Qy       1081 CAGTCAATGACTTATGGATTTAATCAAGGGAAAGTAAATGGCTAACAGTGTTGACAGATCTTCGAAGTTACAGGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CAGTCAATGACTTATGGATTTAATCAAGGGAAAGTAAATGGCTAACAGTGTTGACAGATCTTCGAAGTTACAGGC 1140

Qy       1141 ATGCTAAAAGATTGGGAGGATGACATGCAACAGAAGAGAGTGAAAGAGAGTACTCGGACGA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ATGCTAAAAGATTGGGAGGATGACATGCAACAGAAGAGAGTGAAAGAGAGTACTCGGACGA 1200

Qy       1201 GGAGAAGAAAGAGATCCAGAAGTGAACTAGAACACCAACAAATGTTTAGCAGTATTCAGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGAGAAGAAAGAGATCCAGAAGTGAACTAGAACACCAACAAATGTTTAGCAGTATTCAGC 1260

Qy       1261 AGAGTGAAATTTACTCGTGTGTTACTGACAGTGCTTATAGCCTTTAGCGGTATTTGAAGAGAGACC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGAGTGAAATTTACTCGTGTGTTACTGACAGTGCTTATAGCCTTTAGCGGTATTTGAAGAGAGACC 1320

Qy       1321 AGTGCTGTTGCAGAAGCTCAAAAAATTGATGGTTCAAGCAGCAGCGAGATCTTCTTTCTGCCATT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AGTGCTGTTGCAGAAGCTCAAAAAATTGATGGTTCAAGCAGCAGCGAGATCTTCTTTCTGCCATT 1380

Qy       1381 CATAATTCATTGCATCATGGCATCCAGGCGCCGAATGATGACTAACAAAAGGAGAGATCATCCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 CATAATTCATTGCATCATGGCATCCAGGCGCCGAATGATGACTAACAAAAGGAGAGATCATCCA 1440

Qy       1441 ATTATGATGGGTTTTTGAACCCTTGTGAACCAGAGGCTACTTCCACTACCTTCCGA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 ATTATGATGGGTTTTTGAACCCTTGTGAACCAGAGGCTACTTCCACTACCTTCCGA 1500

Qy       1501 TATGCAAAATAATTAAAAGGGAAGAAAATGGTGAACTACTTATTGTATCCTGGATTTTTTTTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TATGCAAAATAATTAAAAGGGAAGAAAATGGTGAACTACTTATTGTATCCTGGATTTTTTTTC 1560

Qy       1561 ATAAAAAACTGTCTGTGAAGTTTGTGAATTAACAAATTACATTGTATCCCTGGATTTTTTC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 ATAAAAAACTGTCTGTGAAGTTTGTGAATTAACAAATTACATTGTATCCCTGGATTTTTTC 1620

Qy       1621 TGTGAATTTAGTGAACAGTCACCATGTTCTTCTTTCGAAGATCTCGTTACAAAACCACTTTTC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 TGTGAATTTAGTGAACAGTCACCATGTTCTTCTTTCGAAGATCTCGTTACAAAACCACTTTTC 1680

Qy       1681 CTGGTGGATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAAGACACATGGGTGAAAAGATGCA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CTGGTGGATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAAGACACATGGGTGAAAAGATGCA 1740

Qy       1741 CTTCCGGTCTTTTTGTCAGATCCTCCCCAAGTGCCTAATAATTAAATC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CTTCCGGTCTTTTTGTCAGATCCTCCCCAAGTGCCTAATAATTAAATC 1800
```

```
Qy    1741 CTTCCGGTCTTTTGTCAGATCCTCCGAGTGCTTCCCCAAGTGCTACCTAATAATAATC 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db
Qy    1801 ACCAGGCTAAGGACTGTAATCGACTCCCTCCTTGTTGCTGTTCGCCATTCGTAGTC 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 ACCAGGCTAAGGACTGTAATCGACTCCTTGTTGTTACTGTGTTCGCCATTCGTAGTC 1860
Qy    1861 TTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTG 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 TTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTG 1920
Qy    1921 AGGAATTTGCCACTTGCACGGATGAGTTTATGACATTTTATTTATTTAATAGGCAGAGAGGT 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 AGGAATTTGCCACTTGCACGGATGAGTTTATGACATTTTATTTATTTAATAGGCAGAGAGGT 1980
Qy    1981 TGATGCAGGCGCTTCACACCTTGCAGGATGAGTTTGAAACAGGAACCCCAAAGGCCTG 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 TGATGCAGGCGCTTCACACCTTGCAGGATGAGTTTGAAACAGGAACCCCAAAGGCCTG 2040
Qy    2041 TTTAGGTACCTGGGGTCCTTTACCATAACCTTCGCCATTATTGATACAGTACCTTCTCTAAGTGG 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 TTTAGGTACCTGGGGTCCTTTACCATAACCTTCGCCATTATTGATACAGTACCTTCTCTAAGTGG 2100
Qy    2101 CTTTGAATTGGAACTCTACAGTATGCACGGAGTACTATTACATTACATATATTTCTCTCTGA 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 CTTTGAATTGGAACTCTACAGTATGCACGGAGTACTATTACATTACATATATTTCTCTCTGA 2160
Qy    2161 ATTCCTTTACCGCATGGTTGATGTCAACATTGAGTCGTCCGATGGCTCTCAAATGGCAGA 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 ATTCCTTTACCGCATGGTTGATGTCAACATTGAGTCGTCCGATGGCTCTCAAATGGCAGA 2220
Qy    2221 GGAAAGGAATAATGGAAGAGAGCAGCAGAGAAGGCCGTGAGTAGTTGTAAAACAAAGAAAA 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 GGAAAGGAATAATGGAAGAGAGCAGCAGAGAAGGCCGTGAGTAGTTGTAAAACAAAGAAAA 2280
Qy    2281 GAAAGTTGCCCATTGAGCGCGGAGGATCACAAATGAGCCAAGCATATCAGAACATGTGTGC 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 GAAAGTTGCCCATTGAGCGCGGAGGATCACAAATGAGCCAAGCATATCAGAACATGTGTGC 2340
Qy    2341 TGGAATGTTTAAAACCATGGTAGCATTGACATGGACGGGCAAAAGTACGAAGTT 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 TGGAATGTTTAAAACCATGGTAGCATTGACATGGACGGGCAAAAGTACGAAGTT 2400
Qy    2401 TGAGCTTGATAGTGAAACAAGTTCGGTTGTGATGAACACAGGTTTGCTTGCTCCATTCAAACAGTGAT 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 TGAGCTTGATAGTGAAACAAGTTCGGTTGTGATGAACACAGGTTTGCTTGCTCCATTCAAACAGTGAT 2460
Qy    2461 GACCCCGCCGCCCAGTGCCACTACTTACAGTTCAAGGAAATGTCTGACCCCAATAAATATAG 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 GACCCCGCCGCCCAGTGCCACTACTTACAGTTCAAGGAAATGTCTGACCCCAATAAATATAG 2520
Qy    2521 CCCTTCTCCTCAGTGTCTGAACTGTGATGTGGGCAGCTAAGGACACTTTCAACACAGGCAAA 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 CCCTTCTCCTCAGTGTCTGAACTGTGATGTGGGCAGCTAAGGACACTTTCAACACAGGCAAA 2580
Qy    2581 AATGATAATTGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTTAAGGTTGC 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 AATGATAATTGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTTAAGGTTGC 2640
Qy    2641 CAAACCCAACTTTGTGGTTATGGAAGTTATGGCAGGAGGACACAAAAAGGAATCTAAAGT 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 CAAACCCAACTTTGTGGTTATGGAAGTTATGGCAGGAGGACACAAAAAGGAATCTAAAGT 2700
Qy    2701 TCCTCCTGAATTGAATTTCTCTCGCTCATAAATATATTTCCTGTTGTGTTGTGAAACTTGTTTGCAGA 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 TCCTCCTGAATTGAATTTCTCTCGCTCATAAATATATTTCCTGTTGTGTTGTGAAACTTGTTTGCAGA 2760
Qy    2761 GAGACTGGGGAGGTGGCCATAAAGGGGCAGAGTGTTCTTCTTCAGACCCCAACTCTTAGAGGG 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 GAGACTGGGGAGGTGGCCATAAAGGGGCAGAGTGTTCTTCTTCAGACCCCAACTCTTAGAGGG 2820
Qy    2821 CACATCACCAGGCTCCACATCACGGGAGGAAGTGAGATGGAGATGGATTTCTTGGGTAACAACTCATTA 2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 CACATCACCAGGCTCCACATCACGGGAGGAAGTGAGATGGAGATGGATTTCTTGGGTAACAACTCATTA 2880
```

```
Db    2881 TAAGGAATACTTTAGTTGACAGCCTTATAGCCATGATGAAGAAAACTGCTGTTTAAAG 2940
Qy    2881 TAAGGAATACTTTAGTTGACAGCCTTATAGCCATGATGAAGAAAACTGCTGTTTAAAG 2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2941 TGGTTTATTATGTTCCATGGAAGAAAACTGGGTCTTATTGATGAAGTCATTGATGAAGCGTATAT 3000
Qy    2941 TGGTTTATTATGTTCCATGGAAGAAAACTGGGTCTTATTGATGAAGTCATTGATGAAGCGTATAT 3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001 GGTTTTATTACAGATTTAATCACAAAATCATTTTTTATGAATGATGAAGTGAAAATAGTGT 3060
Qy    3001 GGTTTTATTACAGATTTAATCACAAAATCATTTTTTATGAATGATGAAGTGAAAATAGTGT 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3061 TTATAAAGGTTAATAAAATTTCTTGACAAAAAAAAAA 3096
Qy    3061 TTATAAAGGTTAATAAAATTTCTTGACAAAAAAAAAA 3096
```

RESULT 2
ABV23940/c
ID   ABV23940 standard; cDNA; 2668 BP.
XX
AC   ABV23940;
XX
DT   16-SEP-2002 (first entry)
XX
DE   Human prostate expression marker cDNA 23931.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R, Endege WO, Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule and correlating with presence of prostate cancer, useful
PT   prostate cells and correlating with presence of prostate cancer, stage of prostate cancer -
PT   for detecting presence of prostate cancer.
XX
PS   Claim 1; Page 4423; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising
a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
specification or its complement. (I) is useful for:
(a) assessing whether a patient is afflicted with prostate cancer;
(b) monitoring the progression of prostate cancer in a patient;
(c) assessing the efficacy of a test compound to inhibit prostate
cancer in a patient;
(d) assessing the efficacy of a therapy for inhibiting prostate cancer
in a patient;
(e) selecting a composition for inhibiting prostate cancer in a patient;
(f) assessing the prostate cell carcinogenic potential of a compound;
(g) determining whether prostate cancer has metastasized in a patient;
(h) assessing the aggressiveness or indolence of prostate cancer in a
patient;

```
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 2668 BP; 742 A; 617 C; 483 G; 820 T; 6 other;

Query Match        98.8%;    Score 2577.8;   DB 23;   Length 2668;
Best Local Similarity   83.3%;    Pred. No. 0;
Matches 2638;  Conservative 0;  Mismatches 7;  Indels 26;  Gaps 3;


Qy    426 CGCATGCGTGCACCCTGCCGATCGCTCGGGTGGCTGAGAGGGGGGGGGCCGA  485
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2668 CGCATGCGTGCACCCTGCCGATCGCTCGGGTGGCTGAGAGGGGGGGGGCCGA 2609

Qy    486 GGCGGCGGTCGTATTCCGTGGTCCGGACAGCAGTGGGGGGGGGGGGGGGGA  545
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2608 GGCGGCGGTCGTATTCCGTGGTCCGGACAGCAGTGGGGGGGGGGGGGGGGA 2549

Qy    546 AGTAGGCATAAATGGTTTATGAAAGCTTCTGTAGATGATGACGATTCAG    605
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2548 AGTAGGCATAAATGGTTTATGAAAGCTTCTGTAGATGATGACGATTCAG   2489

Qy    606 TATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCA  665
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2488 AGAAGCTTGTCCAGAATTAAAGTTGGGAGAACTACTCATTCATGATAACAA 2429

Qy    666 AGAAGCTTGTCCAGAATTAAAGTTGGGAGAACTACTCATTCATGATAACAA  725
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2428 TGAAGCCAATGTCTGCTCTATTGAAAAATGATGGATCCCAAGATTTTGAAA 2369

Qy    726 TGAAGCCAATGTCTGCTCTATTGAAAAATGATGGATCCCAAGATTTTGAAA  785
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2368 TGAAGCCAATGTCTGCTCTATTGAAATGGATGAGTGCTGGCCATGATTGAA 2309

Qy    786 CCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGAATTAAAT  845
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2308 CCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGAATTAAAT 2249

Qy    846 TAAAGATCTCACCTTGCCTGAACTGATAGGGATTATCATGTTTTTGCTGTT  905
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2248 TAAAGATCTCACCTTGCCTGAACTGATAGGGATTATCATGTTTTTGCTGTT 2189

Qy    906 AACGTGGTTAGAAGCCATTCACACTGGCACAGTAGTTTTACGTTCACATTC  965
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2188 AACGTGGTTAGAAGCCATTCACACTGGCACAGTAGTTTTACGTTCACATTC 2129

Qy    966 TCCAGACTTTATAGAGAAGATCCTGCTATGGAAGGCTTTTGCTCTGGGAAT 1025
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2128 TCCAGACTTTATAGAGAAGATCCTGCTATGGAAGGCTTTTGCTCTGGGAAT 2069

Qy   1026 TGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTCTGTTTTTACATTCA 1085
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2068 TGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTCTGTTTTTACATTCA 2009

Qy   1086 AATGACTTATCGGATTTTAAAATGGCTAACAGTGCAACAGGCATGCT     1145
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   2008 AATGACTTATCGGATTTTAAAATGGCTAACAGTGCAACAGGCATGCT     1949

Qy   1146 AAAAGATTGGGAGGGATGACATGCAAAGAAGAAGAGTACTCGAGTCGACAA 1205
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1948 AAAAGATTGGGAGGGATGACATGCAAAGAAGAAGAGTACTCGAGTCGACAA 1889

Qy   1206 AGAAAGAGATCCAGAAGTTGAACTAGAACCAACCAAGTTTAGCAGTAGT   1265
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1888 AGAAAGAGATCCAGAAGTTGAACTAGAACACAAGTTTAGCAGTAGT     1829

Qy   1266 GAAATTTACTCGTGTGTTACTGACAGTGCCTTTATAGCCTTTATTAAGA   1325
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1828 GAAATTTACTCGTGTGTTACTGACAGTGC                      1769

Qy   1326 TGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGCAGATCTTC    1385
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1768 TGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGCAGATCTTC    1709

Qy   1386 TTCATTGCATCATGGCATCCAGGCCCAGAAATGATACTACAAAAGAGA   1445
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   1708 TTCATTGCATCATGGCATCCAGGCCCAGAAATGATACTACAAAAGAGA   1649

Qy   1446 GATGGGTTTTGAACCCCTTGTAGCCAGGAGCTACTTCACCTTCCACTTC  1505
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1648 GATGGGTTTTGAACCCCTTGTAGCCAGGAGCTACTTCACCTTCCACTTC  1589

Qy   1506 AAAAATAATTAATAAAAGGGAAGAAAATGGTGAAGAAGAAATTTACATT  1565
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1588 AAAAATAATTAATAAAAGGGAAGAAAATGGTGAAGAAGAAATTTACATT  1529

Qy   1566 AATTTAGTGAACAGTCACCCATGTGTTCTTCGTTACAAAACCACTTCCTG 1625
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1528 AATTTAGTGAACAGTCACCCATGTGTTCTTCGTTACAAAACCACTTCCTG 1469

Qy   1626 ATTTAGTGAACAGTGCCATGTGTTCTTCGTTACAAAACCACTTTCCTGT  1685
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1468 ATTTAGTGAACAGTGCCATGTGTTCTTCGTTACAAAACCACTTTCCTGT  1409

Qy   1686 GGATAACAAAAAAGGTCTTTGGAACTCATCTCATCGTCAAAAGATGCACTTCG 1745
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1408 GGATAACAAAAAAGGTCTTTGGAACTCATCTCATCGTCAAAAGATGCACTTCG 1349

Qy   1746 GTCTTTGTCAGATCCTCCGAGTGCTTTCCCCAAGTGCTCTAGTATAATTCACCAG 1805
          ||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1348 GTCTTTGTCAG--TCCTCCGTTGCTTTCCCCAAGTGCTCTAGTATAATTCACCAG 1291

Qy   1806 GCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTTGTTCGGGCCATTCGTAGTCTTATT 1865
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1290 GCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTTGTTCGGGCCATTCGTAGTCTTATT 1231

Qy   1866 CAGATCCATGGACATAACAGGGCTCGGACAGAGGAGTAAGCTTGGTCATATTCTTGAGGAA 1925
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1230 CAGATCCATGGACATAACAGGGCTCGGACAGAGGAGTAAGCTTGGTCATATTCTTGAGGAA 1171

Qy   1926 TTTGCCACCTTGCAGGATGAGTTTATGACACATTTTATTTAAATAGGCAGAGAAGGTTGATG 1985
          ||||||||||||||||||||||||||       ||||||||||||||||
Db   1170 TTTGCCACCTTGCAGGATG--------------AGGCAGAGAAGGTTGATG 1134

Qy   1986 CAGCGCTTCACCACCATGGCTGTTGAAACAGGAACCCCAAAGGCAACACATTTGGCCTGTTTAG 2045
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133 CAGCGCTTCACCACCATGGCTGTTGAAACAGGAACCCCAAAGGCAACACATTTGGCCTGTTTAG 1074

Qy   2046 GTACCTGGGTCCTTTACCATAACCTTCGGCATTATGATACAGTACTTCTAAGTGGCCTTTG 2105
          ||||||||||||||||||||||||||||  |||||||||||||||||||
Db   1073 GTACCTGGGTCCTTTACCATAACCTTCGGCATTATGATACAGTACTTCTAAGTGGCCTTTG 1014

Qy   2106 AATTGGAACTCTACAGTATGCAGGAG-TACTATTACATATTGGTATCTCTCTGAATTC 2164
          |||||||||||||||||||||||||| |||||||||||||||||||||
Db   1013 AATTGGAACTCTACAGTATGCAGGAGTTACTATTACATATTGGTATCTCTCTGAATTC  954

Qy   2165 CTTTACGCATGGTTGATGTCAACATTGAGTCGTCGTGCGATGGCTCTCAAATGGCAGAGGAA 2224
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    953 CTTTACGCATGGTTGATGTCAACATTGAGTCGTCGTGCGATGGCTCTCAAATGGCAGAGGAA  894

Qy   2225 AGGATAATGGAAGAGCAGCAGAAGAAAGGCCGTTAGTAGTAAAAAACAAAGAGAAAGAAA 2284
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    893 AGGATAATGGAAGAGCAGCAGAAGAAAGGCCGTTAGTAGTAAAAAACAAAGAGAAAGAAA  834

Qy   2285 GTTCGCCCCATTGAGCCGGAGAGATCACAATGAGCCAAGCATATCAGAACATGTGTCTGGA 2344
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    833 GTTCGCCCCATTGAGCCGGAGAGATCACAATGAGCCAAGCATATCAGAACATGTGTCTGGA  774

Qy   2345 ATGTTTAAAACCAATGGTAGCATTGACATGGACGCGAAAGTACGTAACCGAAGTTTGAG   2404
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    773 ATGTTTAAAACCAATGGTAGCATTGACATGGACGCGAAAGTACGTAACCGAAGTTTGAG    714

Qy   2405 CTTGATAGTGAACAGTTCGGTATGAACAGGTTTGCTCCATTCAACAGTGTCATGAGACC 2464
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    713 CTTGATAGTGAACAGTTCGGTATGAACAGGTTTGCTCCATTCAACAGTGTGATGACC    654

Qy   2465 CCGCGCGCCAGTGCACTACTTACAGTTCAAGGAGAAAATGTCTGACCTCAAATAAATATAGCCCT 2524
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2465 ...
```

```
Db    653 CCGCGCCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAGCCCT  594
Qy   2525 CCTCCTCAGTTCCTGAACTGTATGTGCAGCTAGTAAGCACTTTCAACACGGCAAAAATG 2584
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    593 CCTCCTCAGTCCCTGAACTGTATGTGCAGCTAGTAAGCACTTTCAACACGGCAAAAATG  534
Qy   2585 ATATTGGAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTAAAGGTTGCCAAA 2644
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    533 ATATTGGAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTAAAGGTTGCCAAA  474
Qy   2645 CCCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAAATCTAAAGTTCCT 2704
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473 CCCAACTTTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAAATCTAAAGTTCCT  414
Qy   2705 CCTGAATTTGATTTCTCTGCCTCATAAATATTTTCCTGTTGTTGAAACTTGTTTGAGAGA 2764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    413 CCTGAATTTGATTTCTCTGCCTCATAAATATTTTCCTGTTGTTGAAACTTGTTTGAGAGA  354
Qy   2765 CTGGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTCAGACCCAACTCTTAGAGGGCACA 2824
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    353 CTGGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTCAGACCCAACTCTTAGAGGGCACA  294
Qy   2825 TCACCAGGCCTCACATCACGGGAAGTGAGATGAGAATTTCTTCTTGGGTAACAACTCATTATAAG 2884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    293 TCACCAGGCTCCACATCACGGGAAGTGAGATGAGAATTTCTTCTTGGGTAACAACTCATTATAAG  234
Qy   2885 GAATACTTTTAGTTTGACAGCCTTATATGATGAATGAAATACTGCTGTTTTAAAGTGGT 2944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    233 GAATACTTTTAGTTTGACAGCCTTATATGACAATGAAATACTGCTGTTTTAAAGTGGT  174
Qy   2945 TTATTATAGTTGTTCCATGGAAGAAACTGGTCTTATTGATTGAATGAAATGCATTGAAGAACGTTATATATGGTT 3004
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    173 TTATTATGTTCCATGGAAGAAACTGGTCTTATTGATTGAATGCATTGAAGAACGTTATAGGTT  114
Qy   3005 TTATTACAGATTTAATCACACAAATCATTTTTTTATGAAATAATAGTTGTTTAT 3064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113 TTATTACAGATTTAATCACAATCATTTTTTTATGAAATAATAGTTGTTTAT   54
Qy   3065 AAAGGTTAATAAATTCTTGACAAAAAAAA 3095
          |||||||||||||||||||||||| ||||
Db     53 AAAGGTTAATAAATTCTTGACAAAACAAA   23
```

RESULT 3
ABV29823/c
ID   ABV29823 standard; cDNA; 2668 BP.
XX
AC   ABV29823;
XX
DT   16-SEP-2002 (first entry)
XX
DE   Human prostate expression marker cDNA 29814.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 200US-183319P.
PR   16-MAR-2000; 200US-189862P.
PR   25-MAY-2000; 200US-207454P.
PR   09-JUN-2000; 200US-211314P.
PR   18-JUL-2000; 200US-219007P.
PR   13-DEC-2000; 200US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX

PI   Schlegel R, Endege WO, Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS   Claim 1; Page 6420; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasied in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacogynamic or pharmacogenomic marker.
XX
SQ   Sequence 2668 BP; 742 A; 617 C; 483 G; 820 T; 6 other;

```
Query Match             83.3%;  Score 2577.8;  DB 23;  Length 2668;
Best Local Similarity   98.8%;  Pred. No. 0;
Matches 2638;  Conservative    0;  Mismatches    7;  Indels   26;  Gaps    3;

Qy   426 CGCATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCAAGATTTGA  485
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2668 CGCATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCAAGATTTGA 2609
Qy   486 GGCGGCGTCGTTATTTCCGTGGTCCGGACAGTGCGTGGCGGGGGGGGGGCCACGGGAGA  545
         |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  2608 GGCGGCGTCGTTATTTCCGTGGTCCGGACAGTGCGTGGCGGGGGGGGGGGGCCACGGGAGA 2549
Qy   546 AGTAGGCATAAATGGTTATGAAAAGCTTCGTGTAGGATGATGACGATTCAGGAGCTCAG  605
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2548 AGTAGGCATAAATGGTTATGAAAAGCTTCGTGTAGGATGATGACGATTCAGGAGCTCAG 2489
Qy   606 TATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCAAGATTTGA  665
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2488 TATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTACCCAAGATTTGA 2429
Qy   666 AGRAGCTTGTCTCGGAGRAATTAAAGTTGGGGAGAACTACTTCATGATGAAGCTATTTGG  725
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2428 AGRAGCTTGTCTCGGAGRAATTAAAGTTGGGGAGAACTACTTCATGATGAAGCTATTTGG 2369
Qy   726 TGRAGCCATGTCTCGCTRATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATGGAAA  785
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2368 TGRAGCCATGTCTCGCTRATTGAAATGATGGATCCCAAGATGGATGCTGGCATGATGGAAA 2309
Qy   786 CCRAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCCAAGGACTGCACTATTAAAAT  845
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2308 CCRAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCCAAGGACTGCACTATTAAAAT 2249
Qy   846 TAAAGATCTCACCTTGCCTGAACTGATAGGGATTATGGATACAATGTTTTGCTGTTTGAT  905
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2248 TAAAGATCTCACCTTGCCTGAACTGATAGGGATTATGGATACAATGTTTTGCTGTTTGAT 2189
Qy   906 AACGTGGTTAGAAGGCCAATCACTGGCACAGACAGTATTTACGTGCCTTTTACATTCATAA  965
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2188 AACGTGGTTAGAAGGCCAATCACTGGCACAGACAGTATTTACGTGCCTTTTACATTCATAA 2129
Qy   966 TCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAAAATCTG 1025
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2128 TCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAAAATCTG 2069
Qy  1026 TGACATTGCAAGGGAAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGAGAGATTTCAGTC 1085
```

```
DT   10-DEC-2002  (first entry)
XX
DE   Human molecule for disease detection and treatment coding sequence #3.
XX
KW   Human; molecule for disease detection and treatment; MDDT; gene therapy;
KW   cytostatic; antiarteriosclerotic; hepatotropic; anti-HIV; antiallergic;
KW   antiinflammatory; antiasthmatic; cerebroprotective; nootropic;
KW   neuroprotective; antiparkinsonian; cardiant; antianginal; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200270709-A2.
XX
PD   12-SEP-2002.
XX
PF   08-FEB-2002; 2002WO-US037709.
XX
PR   09-FEB-2001; 2001US-2681127P.
PR   15-FEB-2001; 2001US-2699618P.
PR   23-FEB-2001; 2001US-2711187.
PR   07-MAR-2001; 2001US-2744867.
PR   09-MAR-2001; 2001US-2744367P.
PR   28-NOV-2001; 2001US-3342297P.
PR   01-FEB-2002; 2002US-3532842P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lal PG, Baughn MR, Yao MG, Walla NK, Elliot VS, Xu Y;
PI   Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DAM;
PI   Hafala AJA, Ghandi AR, Thangavelu K, Sanjanwala MM, Tang YT;
PI   Ramkumar J, Griffin JA, Swarnaker A, Azimzai Y, Sapperstein SK;
PI   Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;
XX
DR   WPI; 2002-713453/77.
DR   P-PSDB; AAO19400.
XX
PT   New human molecules for disease detection and treatment (MDDT), useful
PT   for diagnosing, treating and preventing diseases or conditions
PT   associated with the aberrant MDDT expression, e.g. cancer, AIDS,
PT   asthma, diabetes, hepatitis -
XX
PS   Claim 11; Page 159-160; 177pp; English.
XX
CC   The present invention relates to human proteins and coding sequences of
CC   molecules for disease detection and treatment MDDT. The sequences can be
CC   used in the treatment of diseases associated with the decreased
CC   expression or overexpression of MDDT, such as cell proliferative (cancer,
CC   atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS, renal
CC   allergies, Addison's disease, asthma), developmental (dwarfism, renal
CC   tubular acidosis), neurological (e.g. stroke, Parkinson's disease,
CC   epilepsy) and cardiovascular (congestive heart failure, myocardial
CC   infarction, angina pectoris) disorders. The present sequence is a
CC   coding sequence of the invention.
XX
SQ   Sequence 2488 BP; 768 A; 461 C; 567 G; 692 T; 0 other;
```

```
Query Match             78.1%;  Score 2417.2;  DB 24;  Length 2488;
Best Local Similarity   98.5%;  Pred. No. 0;  Matches 2470;  Conservative   0;  Mismatches  13;  Indels  25;  Gaps  2;
```

```
Qy      474 GGGCGGGCGGGCCGAGGGCGGCGTCGTTATTTCCGTGGTCCGGACAGTGCGTGGGGCCGGGGT 533
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GGCGGCGGCGCGCGAGCCGCGGCGCGCGTCGTTATTTCCGTGGTCCGGACAGTGCGTGGGGCCGGGGT 60

Qy      534 GACCACGCGGGAGGAGGTAGGCGCATAAATGGTTATAGAAGCTTCTGTAGATGATGACGATTCAGG 593
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GACCACGGGGAGGAGAAGTAGGCATAAATGGTTATAGAAAGCTTCTGTAGATGATGACGATTCAGG 120

Qy      594 ATGGGAGCTCAGTATGCCAGAAAAAAATGGAGAAAAAGCAATAACTACAAACTGGGTGGACATTAC 653
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ATGGGAGCTCAGTATGCCAGAAAAAAATGGAGAAAAAGCAATAACTACAAACTGGGTGGACATTAC 180

Qy      654 CCAAGATTTTGAAGAGATTTAAAGCTTGTCGGAGAACTACTTCAGTGATGATAAGGCT 713
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 CCAAGATTTTGAAGAAGAGCTTGTTGTCGGAGAATTAAAGTTGGGAGAACTACTTCATGATGATAAGGCT 240

Db      241 ATTTGGTCTTTTTGAAGCCATGGGCATGCTTGCTATTGATGAAATGATTGGATGCTGG 300

Qy      714 ATTTGGTCTTTTTGAAGCCATGGGTGCTTGCTATTGATGAAATGATTGGATGCTGG 773
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 CATGATTGGAAACCAAGTTAATCGAAAAAGTTCTCAATTTTGAAGCAAGGATGG 360

Qy      774 CATGATTGGAAACCCAAGTTAATGAAAAAGTTCTCAATTTTGAAGCAAGGATGG 833
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CACTATTAAAAAATTAAAGATCTCACCTTGCCTGAAACTGATAGGGATTATGGATACATGTTT 420

Qy      834 CACTATTAAAAAATTAAAGATCTCACCTTGCCTGAAACTGATAGGGATTATGGATACATGTTT 893
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 TTGCTGTTTGATAACGTGGTTAGAAGGCCATTCACTCGACACAGTATTTACGTGCCT 480

Qy      894 TTGCTGTTTGATAACGTGGTTAGAAGGCCATTCACTCGACACAGTATTTACGTGCCT 953
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 TTACATTCATAATCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTGCTCTGGGAAT 540

Qy      954 TTACATTCATAATCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTGCTCTGGGAAT 1013
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 CTTGAAAAATCTGTGACATTGCAAGGGGAAAAAGTAAAATAAAGCTCGTGTTTTTGAAGAGA 600

Qy     1014 CTTGAAAAATCTGTGACATTGCAAGGGGAAAAAGTAAAATAAAGCTCGTGTTTTTGAAGAGA 1073
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 TACAGGCCATGCTAAAAGATGTGGAGGATGACACAGTGAAAGAGAGAGTACTCGGAAG 660

Qy     1074 AGATTTTCAGTCAATGACTCTATGGATTTTAAAATGGCTAACAGTGTGACAGATCTTCGAGT 1133
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 TACAGGCCATGCTAAAAGATGTGGAGGATGACACAGTGAAAGAGAGAGTACTCGGAAG 720

Qy     1134 TACAGGCCATGCTAAAAGATGTGGAGGATGACACAGTGAAAGAGAGAGTACTCGGAAG 1193
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 TCGACAAGGAGAAGAAGAAGATCCAGAAGTTGAACTAGAACACCAACAATGTTAGCAGT 780

Qy     1194 TCGACAAGGAGAAGAAGAAGAAGATCCAGAAGTTGAACTAGAACCAACAATGTTAGCAGT 1253
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 ATTCAGCAGAGTGAAAATTTACTCCTGTGTTACTGACAGTGCTTTACTACTAAGAA 840

Qy     1254 ATTCAGCAGAGTGAAAATTTACTCCTGTGTTACTGACAGTGCTTTACTACTAAGAA 1313
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 AGAGACAGCAGTGCTGTTGCAGAAGCTCAAAAATTGATTGGTTCAAGCAGCAGATCTTCTTTC 900

Qy     1314 AGAGACCAGTGCTGTTGCAGAAGCTCAAAAATTGATTGGTTCAAGCAGCAGATCTTCTTC 1373
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 TGCCATTCATAATTCATCATGCATGCATCCAGCCCCAGAATGATACTACAAAAGGAGA 960

Qy     1374 TGCCATTCATAATTCATCATGCATGCATCCAGCCCCAGAATGATACTACAAAAGGAGA 1433
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 TCATCCAATTATGATGGGTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACCTACCTT 1020

Qy     1434 TCATCCAATTATGATGGGTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACCTACCTT 1493
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 CCCTCGATATGCAAAAATAATTAAAAGGGAAGAAAATGGTGAACTATTTTGCAAGATTAAT 1080

Qy     1494 CCCTCGATATGCAAAAATAATTAAAAGGGAAGAAAATGGTGAACTATTTTGCAAGATTAAT 1553
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 AGATAGAATAAAAAACTGTCTGTGAGGTTGTGAATTAACAAAAATTACATTGTATCCTGGA 1140

Qy     1554 AGATAGAATAAAAAAACTGTCTGTGAGGTTGTGAATTAACAAAATTACATTGTATCCTGGA 1613
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 TTTTTTCTGTGAATTAGTGAACAGTCACCAATGGTTCTTTACAAAAC 1200

Qy     1614 TTTTTTCTGTGAATTAGTGAACAGTCACCAATGGTTCTTTACAAAAC 1673
            ||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 CACTTTCCTGGGTAACAAAAAAGGTCTTTGGAACTCATCTCATGGTGAA 1260

Qy     1674 CACTTTCCTGGGTAACAAAAAGGTCTTTGGAACTCATCTCATGGTGAA 1733
            |||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 AGATGCACTTCGGGTTCTTTTGTCGAGATCCTTCCCCAAGTGCTTACCTTATAT 1793

Qy     1734 AGATGCACTTCGGGTTCTTTTGTCGAGATCCTTCCCCAAGTGCTTACCTTATAT 1793
            |||||||||||||||||||||
```

```
Db  1261 AGATGCACTTCGGTCTTTTGTTCGTTCAGTCCTCCG--GTGCTTCCCCCAAGTGCTTACCTATAT 1318
Qy  1794 AATAATCACCAGGCTAAGGACTGTATCGGATCTCCTTTGTTACTCACTGTGTTCGGCCATTC 1853
Db  1319 AATAATCACCAGGCTAAGGACTGTATCGGATCTCCTTTGTTACTCACTGTGTTCGGCCATTC 1378
Qy  1854 TGTAGTCTTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAT 1913
Db  1379 TGTAGTCTTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCAT 1438
Qy  1914 ATTCTTGAGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGCAG 1973
Db  1439 ATTCTTGAGGAATTTGCCACCTTGCAGGATG---------AGGCAG 1475
Qy  1974 AGAAGGTTGATGCAGCGCTTCACACCATGCTTGTTGAAACAGGAACCCCAAAGGCAACATT 2033
Db  1476 AGAAGGTTGATGCAGCGCTTCACACCATGCTTGTTGAAACAGGAACCCCAAAGGCAACATT 1535
Qy  2034 TGGCCTGTTGTTAGGTACCTGGGTCCTTTACCATAACTTCGCATTATGATACAGTACCTTC 2093
Db  1536 TGGCCTGTTGTTAGGTACCTGGGTCCTTTACCATAACTTCGCATTATGATACAGTACCTTC 1595
Qy  2094 TAAGTGGCTTTGAATTGGAACTCTCACGTATGCACGAGTACTATTACATATATTATTGGTATC 2153
Db  1596 TAAGTGGCTTTGAATTGGAACTCTCACGTATGCACGAGTACTATTACATATATTATTGGTATC 1655
Qy  2154 TCTCTGAATTCCTTTACGCATGGTTGAATGTCAACATTGAGTCGTGTCCGATGGCCTCCAAA 2213
Db  1656 TCTCTGAATTCCTTTACGCATGGTTGAATGTCAACATTGAGTCGTGTCCGATGGCCTCTCAAA 1715
Qy  2214 TGGCAGGAGGAAAGGGATAAATGGGAAGAGAGAGCAGAAAGGCCGTAGTAGTAGTAAAAACAAAGA 2273
Db  1716 TGGCAGGAGGAAAGGGATAAATGGGAAGAGAGAGCAGAAAGGCCGTAGTAGTAGTAAAAACAAAGA 1775
Qy  2274 AAAAAAAGAAAAGTTCGCCCATTGAGCCGAGGATCACAATGAGCCAAGCATATCAGAACA 2333
Db  1776 AAAAAAAGAAAAGTTCGCCCATTGAGCCGAGGATCACAATGAGCCAAGCATATCAGAACA 1835
Qy  2334 TGTGTGCTGGGAATGTTTAAAAACCATGGTAGCATTGACATGGACGGCAAAGTACGTAAAC 2393
Db  1836 TGTGTGCTGGGAATGTTTAAAAACCATGGTAGCATTGACATGGACGGCAAAGTACGTAAAC 1895
Qy  2394 CGAAGTTTGAGCTTGATAGTAGTGAACAAGTTCGGTAGTGAACACAGGTTTGCTCCATTCAACA 2453
Db  1896 CGAAGTTTGAGCTTGATAGTAGTGAACAAGTTCGGTAGTGAACACAGGTTTGCTCCATTCAACA 1955
Qy  2454 GTGTGATGAACCCGCGCGCCAGTGCACTACTTTACAGTTCAAGGAAAGTCTGACCTCAATA 2513
Db  1956 GTGTGATGAACCCGCGCGCCAGTGCACTACTTTACAGTTCAAGGAAAGTCTGACCTCAATA 2015
Qy  2514 AATATAGCCCCTCCTCCTCAGTCTCCTCGAACCTGTATGTTGGCACTACACTTTCAAC 2573
Db  2016 AATATAGCCCCTCCTCCTCAGTCTCCTCGAACCTGTATGTTGGCACTACACTTTCAAC 2075
Qy  2574 AGGCAAAAAATGATATTGGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTTAA 2633
Db  2076 AGGCAAAAAATGATATTGGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTTAA 2135
Qy  2634 AGGTTGCCAAACCCACTTGTGGTTATGAAGTTGATTGGCAGGAGGACACAAAAGGAAT 2693
Db  2136 AGGTTGCCAAACCCACTTGTGGTTATGAAGTTGATTGGCAGGAGGACACAAAAGGAAT 2195
Qy  2694 CTAAAGTTCCTCCTGAATTTGAATTTCTGCTCATAAAATATTTTCCTGTTGTGTGAAACTTG 2753
Db  2196 CTAAAGTTCCTCCTGAATTTGAATTTCTGCTCATAAAATATTTTCCTGTTGTGTGAAACTTG 2255
Qy  2754 TTTTGAGAGAGACTGGGGAGGTGGCCATAAAGGAGGCAGAGTCTTCTTTCAGACCCAACTCT 2813
Db  2256 TTTTGAGAGAGACTGGGGAGGTGGCCATAAAGGAGGCAGAGTCTTCTTTCAGACCCAACTCT 2315
Qy  2814 TAGAGGGCACAATCACCAGGCTCCACATCACGGGGAAGTGAGATGGATTTCTTGGGTAACAA 2873
Db  2316 TAGAGGGCACAATCACCAGGCTCCACATCACGGGGAAGTGAGATGGATTTCTTGGGTAACAA 2375

Qy  2874 CTCATTATAAGGAATACTTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGT 2933
Db  2376 CTCATTATAAGGAATACTTTTTAGTTTGACAGCCTTATATGACATGAATGAAAACTGCTGT 2435
Qy  2934 TTTAAAGTGGTTTATTTATGTTCCATGGAAGAGAAAACTGGTCTTATTGAAT 2981
Db  2436 TTTAAAGTGGTTTATTTATGTTCCATGAAGTAAGACACTGGGTTCCATTAAT 2483
```

RESULT 5

ABS51309

ID   ABS51309 standard; cDNA; 719 BP.
XX
AC   ABS51309;
XX
DT   21-OCT-2002 (first entry)
XX
DE   cDNA encoding human secretory protein #7.
XX
KW   Human; secretory polypeptide; SPTM; actinic keratosis; arteriosclerosis;
KW   bursitis; cirrhosis; hepatitis; polycythaemia vera; anaemia; psoriasis;
KW   primary thrombocytopenia disorder; cancer; adenocarcinoma; leukaemia; myeloma;
KW   sarcoma; immune system disorder; acquired immunodeficiency syndrome;
KW   AIDS; allergy; asthma; Crohn's disease; diabetes mellitus; gout;
KW   glomerulonephritis; Goodpasture's syndrome; thyroiditis; pancreatitis;
KW   hepatitis; multiple sclerosis; osteoporosis; Reiter's syndrome;
KW   rheumatoid arthritis; neurological disorder; epilepsy; stroke; dementia;
KW   Alzheimer's disease; Pick's disease; Huntington's disease; mood; anxiety;
KW   Parkinson's disease; central nervous system disorder; mental disorder;
KW   schizophrenic disorder; amnesia; Tourette's disorder; transgenic animal;
KW   gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200257304-A2.
XX
PD   25-JUL-2002.
XX
PF   15-JAN-2002; 2002WO-US01340.
XX
PR   16-JAN-2001; 2001US-261864P.
PR   16-JAN-2001; 2001US-261865P.
PR   16-JAN-2001; 2001US-261979P.
PR   16-JAN-2001; 2001US-261981P.
PR   17-JAN-2001; 2001US-262164P.
PR   17-JAN-2001; 2001US-262208P.
PR   17-JAN-2001; 2001US-263131P.
PR   19-JAN-2001; 2001US-262599P.
PR   19-JAN-2001; 2001US-262760P.
PR   19-JAN-2001; 2001US-263063P.
PR   19-JAN-2001; 2001US-263066P.
PR   19-JAN-2001; 2001US-263069P.
PR   19-JAN-2001; 2001US-263070P.
PR   19-JAN-2001; 2001US-263074P.
PR   19-JAN-2001; 2001US-263076P.
PR   19-JAN-2001; 2001US-263077P.
PR   19-JAN-2001; 2001US-263329P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Panzer SR, Lincoln SE, Altus CM, Dufour GE, Hillman JL, Jones AL;
PI   Dam TC, Liu TF, Harris B, Flores V, Daffo A, Marwaha R, Chen AJ;
PI   Chang SC, Gerstin EH, Peralta CH, David MH, Lewis SA;
XX
DR   WPI; 2002-507016/63.
DR   P-PSDB; ABG65816.
XX
PT   New purified secretory polypeptides and polynucleotides, useful in the
PT   diagnosis, study, prevention or treatment of diseases associated with
PT   decreased expression of functional secretory molecules, e.g. AIDS,
PT   cancer or allergies -
XX

PS   Claim 1; Page 247; 340pp; English.
XX
CC   The invention describes an isolated polynucleotide a naturally occurring
CC   polynucleotide sequence at least 90 % identical to it, a polynucleotide
CC   complementary to it or an RNA equivalent of it. The purified secretory
CC   polypeptides (SPTM) and polynucleotides are useful in the diagnosis,
CC   polypeptides (SPTM) e.g. actinic keratosis, arteriosclerosis,
CC   study, prevention or treatment of diseases associated with decreased
CC   expression of functional SPTM, e.g. actinic keratosis, arteriosclerosis,
CC   bursitis, cirrhosis, hepatitis, polycythaemia vera, primary
CC   thrombocytopenia, anaemia, psoriasis, cancers including adenocarcinoma,
CC   leukaemia, myeloma or sarcoma, immune system disorder such as acquired
CC   immunodeficiency syndrome (AIDS), allergies, asthma, Crohn's disease,
CC   diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout,
CC   Hashimoto's thyroiditis, hepatitis, multiple sclerosis, osteoporosis,
CC   pancreatitis, Reiter's syndrome, autoimmune thyroiditis or rheumatoid
CC   arthritis, neurological disorders such as epilepsy, stroke, Alzheimer's
CC   disease, Pick's disease, Huntington's disease, dementia, Parkinson's
CC   disease, other developmental disorder of the central nervous system,
CC   mental disorder including mood, anxiety or schizophrenic disorder,
CC   amnesia or Tourette's disorder. The polynucleotides may be used in
CC   hybridisation and amplification technologies, e.g. in assessing gene
CC   expression patterns, to develop a transcript image for a particular cell
CC   or tissue, or to create transgenic animals to model human disease. This
CC   sequence encodes a human secretory protein isolated in the invention.
XX
SQ   Sequence 719 BP; 223 A; 113 C; 188 G; 195 T; 0 other;

Query Match   23.2%; Score 717.4; DB 24; Length 719;
Best Local Similarity   99.8%; Pred. No. 7e-169;
Matches 718; Conservative   0; Mismatches 1; Indels   0; Gaps   0;

Qy   474 GGCGGGGCGGCGGCCGGAGGGGGGGTCGTCGTTGTTTTCCGTGGTCCGGACAGTGCCTGGCCGGGGGT 533
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGCGGGGCGGCGGCCGGAGGGGGGGTCGTCGTTGTTTTCCGTGGTCCGGACAGTGCCTGGCCGGGGGT 60

Qy   534 GACCACCGGGAGAAAGTAGGCCAATAATGGTTATGAAAAGCTTCTTGTAGATGATGAACGATTCAGG 593
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GACCACCGGGAGAAAGTAGGCCAATAATGGTTATGAAAAGCTTCTTGTAGATGATGAACGATTCAGG 120

Qy   594 ATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTAC 653
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTAC 180

Qy   654 CCAAGAGTTTGAAGGAGGCTTGTCGACGAATTAAAGTTGGGAGAACTACTTCATGATAAGCT 713
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CCAAGAGTTTGAAGGAGGCTTGTCGACGAATTAAAGTTGGGAGAACTACTTCATGATAAGCT 240

Qy   714 ATTTGGTCTTTTTGAAGCCATGTCTCGCTATTGAAAATGATGAGATCCCAAGATGGATGCTGG 773
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ATTTGGTCTTTTTGAAGCCATGTCTCGCTATTGAAAATGATGAGATCCCAAGATGGATGCTGG 300

Qy   774 CATGATTGGAAACCAAGGTTAATCGAAAGTTCTCAATTTTGAACAGCTATCAAGGATGG 833
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CATGATTGGAAACCAAGGTTAATCGAAAGTTCTCAATTTTGAACAGCTATCAAGGATGG 360

Qy   834 CACTATTAAAATTAAAGAATCTCACCTTGCCTGACTGATAGGGATTATGGATACATGTTT 893
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CACTATTAAAATTAAAGAATCTCACCTTGCCTGACTGATAGGGATTATGGATACATGTTT 420

Qy   894 TTGCTGTTTGATAACGTGGTTAGAGAGGCCAATTCACTGGCACAGCAGAGTATTTACGTGCCT 953
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TTGCTGTTTGATAACGTGGTTAGAGAGGCCAATTCACTGGCACAGCAGAGTATTTACGTGCCT 480

Qy   954 TTACATTCATAATCCAGACTTTATAGAAGATCCTGCTGCACAGGAGAAAAGTAAATAAAGCTT 1013
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TTACATTCATAATCCAGACTTTATAGAAGATCCTGCTGCACAGGAGAAAAGTAAATAAAGCTT 540

Qy   1014 CTTGAAAATCTGTGACAATTGCACAAGGAAAAAAGTAAATAAAAGCTTGCTGTTTTTGAAGAGGA 1073
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CTTGAAAATCTGTGACAATTGCACAAGGAAAAAAGTAAATAAAAGCTTGCTGTTTTTGAAGAGGA 600

Qy   1074 AGATTTCAGTCAATGACTTATGGATTAAAAATGGCTAACAGTGTGACAGATCTTCGAGT 1133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AGATTTCAGTCAATGACTTATGGATTAAAAATGGCTAACAGTGTGACAGATCTTCGAGT 660

Qy   1134 TACAGGCATGCTAAAAGATGTGGGAGGATGGACAGATGCAAAGAAGAAGTAACAGTACTCGAA 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TACAGGCATGCTAAAAGATGTGGGAGGATGGACAGATGCAAAGAAGAAGTAACAGTACTCGAA 719

RESULT 6
ID   ABX50281
AC   ABX50281;
XX
DT   25-FEB-2003 (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #210.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2021371160-Al.
XX
PD   26-SEP-2002.
XX
PF   26-OCT-2001; 2001US-0983965.
XX
PR   17-DEC-1998; 98US-1136782P.
PR   15-DEC-1999; 99US-0465231.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
PR   WPI; 2003-102386/09.
XX
PT   Purified nucleic acid molecules, useful for genome mapping, gene
PT   identification and analysis, cattle breeding or preparation of
PT   constructs for cattle gene expression and genetically improved cattle -
XX
PS   Claim 2; SEQ ID No 210; 38pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived
CC   from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC   second nucleic acid molecule comprising any of 5912 nucleotide
CC   sequences, appearing as ABX50072-ABX55983, or complements of them.
CC   Also included are; (1) a transformed cell having a nucleic acid
CC   comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC   translated sequence that functions in the cell to cause termination of
CC   transcription and addition of polyadenylated ribonucleotides to a 3' end
CC   of the mRNA molecule; and (2) determining a level or pattern of a
CC   molecule in a bovine cell or tissue comprising: (a) incubating any of the
CC   nucleic acid (comprising any of the 5912 nucleic acid sequences or its
CC   complement or fragment) with a complementary nucleic acid molecule
CC   obtained from the bovine cell or tissue, where hybridisation between the
CC   marker nucleic acid and the complementary nucleic acid permits the
CC   detection of the molecule; and (b) detecting the level or pattern of the
CC   complementary nucleic acid, where the detection of the complementary
CC   nucleic acid is predictive of the level or pattern of the molecule.
CC   The LMFD nucleic acid is used for determining a level or pattern
CC   of a molecule in a bovine cell or tissue. It is useful for genome
CC   mapping, gene identification and analysis, cattle breeding, preparation
CC   of constructs for use in cattle gene expression, or for genetically
CC   improving cattle. The present sequence is one of the 5912 bovine
CC   LMFD EST (expressed sequence tag) nucleic acids.
CC   Note: The present sequence was not shown in the specification but
CC   was obtained in electronic format from the USPTO web site;

Qy   1 ABX50281
              ||||||||
Db   1134 TACAGGCATGCTAAAAGATGTGGGAGGATGGACAGATGCAAAGAAGAAGTAACAGTACTCGAA 1192
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TACAGGCATGCTAAAAGATGTGGGAGGATGGACAGATGCAAAGAAGAAGTAACAGTACTCGAA 719

Db   601 AGATTTCAGTCAATGACTTATGGATTAAAAATGGCTAACAGTGTGACAGATCTTCGAGT 660

```
CC   seqdata.uspto.gov/sequence.html?DocID=2002037160.
XX
SQ   Sequence 398 BP; 131 A; 89 C; 85 G; 93 T; 0 other;

     Query Match     11.3%;  Score 350.8;  DB 25;  Length 398;
     Best Local Similarity   94.3%;  Pred. No. 2.1e-77;
     Matches  364;  Conservative   0;  Mismatches  22;  Indels   0;  Gaps   0;

Qy   2298 GCCCAGAGATCACATAATGAGCCAAGCAATATCAGAACAATGTGTGTGCTGGAATAGTTTAAAACCA 2357
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13 GCCCAGAGATCACAATGAGCCAAGCAATATCAGAACAATGTGTGTGCTGGAATAGTTTAAAACCA 72

Qy   2358 TGGTAGCATTTGACATGGCACGGCCAAGCTAGAAACCCAAAGTTTGAGCTTGATAGTAGTGAAC 2417
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     73 TGGGAGCTTTTGACATGGATGGATTCGAAAACCCAAGTTTGAGCTTGATAGTAGTGAAC 132

Qy   2418 AAGTTCGGTATGAAACACAGGTTTGCTCTCATTCAAACAGTGTGGATGAGACCCCGGCCAGTGC 2477
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    133 AAGTTCGGATATGAGCACAGAATTTGCTCCATTCACAGTGTAATGACACCACCAGTGC 192

Qy   2478 ACTACTTACAGTTCAAGGAAAATGTTCTGACCTCAATAAATATATAGCCCTCCTCCTCCAGTCTC 2537
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193 ACTATCTGCAGTTCAAGGAAAATGTCTGACCTCAATAAATATAGCCCTCCTCCTCCAGTCTC 252

Qy   2538 CTGAACTGTATGTGGCAGCTAGTAAGCACTTTCACAGGCAAAAATGTATTGGGAAATA 2597
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253 CAGAACTGTATGTGGCAGCTAGTAAGCACTTTCAGCAGGCAAAAATGATTACTGGAAAATA 312

Qy   2598 TTCCTAACCCGGACCACATGAGGTTAATAGAATTTAAAGGTTGCCAAACCCAACTTTGTTGTGG 2657
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    313 TCCCAAACCCAGACCATGAGGTCAATAGAATTTAAAGGTTGCCAAACCCAACTTTGTGTGG 372

Qy   2658 TTATGAAGTTATTGGCAGGAGGACAC 2683
          ||||||||||||||||||||||||||
Db    373 TTATGAAGTTATTGGCAGGAGGACAC 398
```

```
RESULT 7
ID   ABQ51658
ID   ABQ51658 standard; DNA; 799 BP.
XX
AC   ABQ51658;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 38249.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
```

```
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers.
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, particularly by detecting mutations or single nucleotide
CC   systems etc., cardiovascular, gastrointestinal and respiratory
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 799 BP; 118 A; 95 C; 297 G; 289 T; 0 other;
```

```
     Query Match     11.1%;  Score 343.8;  DB 24;  Length 799;
     Best Local Similarity   83.5%;  Pred. No. 1.7e-75;
     Matches  390;  Conservative   0;  Mismatches  77;  Indels   0;  Gaps   0;

Qy     87 TGCTCGTTCGTGCTGCTCGGCAACCACTAAGGTCTACGGCAAACCTCCACGGTTTCCTTCCGCCT 146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 TGTTCGTTCGTCGTTCGGTTAATATTAATTAAGGTTACGTAAATTTTACGGTTTTTCGTTT 61

Qy    147 TCGCGTCACCTTTCTAAGAAAATTCCCAGAGGGAGGGCAGCAGCGAGACACGGGGCGGGGCG 206
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 TCGCGTTATTTTGTTTTTTAGGAAAATTTTAGAGGGTAGGCGTAGACGGGGCGGGTTTGAGATT 121

Qy    207 CCGGGGCTCCGCCCTCTTCCCGGGAACCGCCACTACCCAGGAACTCCGACAGAGGGGTGAAAA 266
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 TCGGGTTTCGTTTTTTTGTTTTTTTGGGGAATCGTTTATTATTTAGGATTTCGGATAGAGGGTGAAA 181

Qy    267 AAGATAACTTCCCGGTGCCGGATGCGTCTCTAATCTCGGCGAGGAGGAGGGCGCCGCCAT 326
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 AAGATAATTTCCGGTTCGGATCGTTTTTAATTCGGCGAGAGAGGAGGGCGTTAT 241

Qy    327 CGGCCGAACGGACGCGTGGGGGGCGAGGAGGGGGAGTGTGGCCGGGAGGGCCGAAGTCCCCGGG 386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 CGGTCGAACGGACGCGTGGCGAGGAGGGGGATACGGCGATTACGGAGGGGGGAAGTTTCGG 301

Qy    387 AGTAAGGGAGGAGGGGGGGGGGTCGGCGGTCGGGGGCGATACGGATACGGACGCGTGCCGG 446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 AGTAAGGGAGGAGGGGGGTCGGCGGTTTCGGGTATACGTATGCGTGTACGGTTCGG 361

Qy    447 TCGGGCTCGGGCTGAGAGGGGAGGGGGGGGGGGGGGCCCCAGGGCGGCGCGTCGTTATTTCGGTG 506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 TCGGGTTGGGTTGAGAGGGGAGGGGGGGGGGGGGTCGAGGCGGCGTCGTTATTTCGGTG 421

Qy    507 GTCCCGGACAGTGCGTGGCGTGGCGCGGGGTTGCCACGGGGAGGAAGTAGGCA 553
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 GTTCGGATAGTGCGTGGCGCGGGGTGATTACGGGAGGAAGTAGGTA 468
```

```
RESULT 8
ID   ABQ51659/c
ID   ABQ51659 standard; DNA; 799 BP.
XX
AC   ABQ51659;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 38250.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
```

KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 799 BP; 289 A; 297 C; 95 G; 118 T; 0 other;

Query Match   11.1%; Score 343.8; DB 24; Length 799;
Best Local Similarity   83.5%; Pred. No. 1.7e-75;
Matches 390; Conservative   0; Mismatches   77; Indels   0; Gaps   0;

Qy   87 TGCTCGTTCGTTCGTCGCAACCACTAAGGTCTACGGAAACCTCACGGTTCCTTCCGCCT 146
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   798 TGTTCGTTCGTGTTCGTAATTATTAAGGTTACGTAAAATTTTACGGTTTTTTTCGTTT 739

Qy   147 TCGGCGTCACCTTTCTAAGAAAATTCCCAGAGGCAGCGCAGGGCGGGGGCTCTGAGACT 206
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   738 TCGCGTTATTTTTTTTAAGAAATTTTTAGAGGGTAGCGTAGGACGGGGGGTTTGAGATT 679

Qy   207 CCGGGCTCCGCCTCTTCTTCCGGAACCGCCACTACCCAGACTCCGACAGAGGGTGAAAA 266
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   678 TCGGGTTTCGTTTTTTTCGGGAATCGTTTATTTATTTAGGATTTCGATAGAGGGTGAAA 619

Qy   267 AAGATAACTTCCGGTCTCCGCGATGTGTCTCTAATCTCGCGAGAGAAGAGAGGGCCGCCAT 326
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   618 AAGATAATTTCGGTTTCGGCGATCGTTTTTTAATTCGCGAGAAGAGAGAGGGTTTAT 559

Qy   327 CGGCCGAACGGAGCGGGTTGCGAGGGAGGGGGGGGTGTGGCGGGCGAAGTCCCCGGG 386
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   558 CGGTCGAACGGAGCGGTGCGGGGAGGGGGGGTGTGGTCGGGGGAGGACGCGGAAGTTTTCGGG 499

Qy   387 AGTAAGGAGAGAGGGCGGCGCGCCGGGGGGGGTGGCGCGTCCCGGGCGCATACGGAACGCTGCCGG 446
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   498 AGTAAGGAGAGGGGCGGCGGCGGTCGGCGGGGGGGGGTTTCGGGGTTATACGTATTACGTATGGTGCG 439

Qy   447 TCGGGGCGTGGGGCTGAGAGGGGGAGGGGGAGGGGGGGGCGCGCCGAGGGTCGGTTGTTATTTCCGTG 506
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   438 TCGGGTTGGGGTTGAGAGGAGGGAGGCGGGGCGCGGGGGGGGGGTCGGAGGGGGCGGTCGTTATTTTCGTTG 379

Qy   507 GTCCGGACAGTGCGTGGGCGCGGGCGGGGTGACCACGGGAGAAGTAGGCA 553
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   378 GTTCGGATAGTGCCTGGGCTGGCGCGGGGGTGATTACGGGAGAAGTAGGTA 332

RESULT 9
ID ABV15331/c
AC ABV15331 standard; cDNA; 356 BP.
XX
DT 13-SEP-2002 (first entry)
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
OS Homo sapiens.
XX
DE Human prostate expression marker cDNA 15322.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-2074454P.
PR 09-JUN-2000; 2000US-2113314P.
PR 18-JUL-2000; 2000US-2190079P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 2572; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (II) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 356 BP; 97 A; 80 C; 85 G; 94 T; 0 other;

Query Match   9.4%; Score 292.4; DB 23; Length 356;

Best Local Similarity   90.5%; Pred. No. 7.7e-63;
Matches 344;   Conservative   0; Mismatches 11; Indels 25; Gaps   2;

Qy   1671 AACACACTTTCCTCGGTGGGTGGGATAACAAAAAGGTCTTTGGAACTCATCTCATGGCAAGACATGGT 1730
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   355 AAACCACTTTCCGGTGGGATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGT 296

Qy   1731 GAAAGATGCACTTCGGTCTTTTGTCAGATCCTCCGAGTGGCTTTCCCCAAGTGCTACCTA 1790
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   295 GAAAGATGCACTTCGGTCTTTTGTCAG--TCCTCCGGGTGCTTTCCCCAAGTGCTACCTA 238

Qy   1791 TATAATAATCACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTTCGGCCA 1850
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237 TATAATAATCACCAGGCTAAGGACTGTATCGACTCCTTTGTTACTCACTGTTCGGCCA 178

Qy   1851 TTCTGTAGTCTTATTCAGATCCATGGACATAACCAGGGCTCGGACAGAGAGATAAGCTTGGT 1910
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177 TTCTGTAGTCTTATTCAGATCCATGGACATAACCAGGGCTCGGACAGAGAGATAAGCTTGGT 118

Qy   1911 CATATTCTTGAGGAATTTGCCACCTTGCAGGATGAGTTTATGACATTTATTTTAATTAATAGG 1970
          ||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   117 CATATTCTTGAGGAATTTGCCACCTTGCAGGATG--------------AGG 81

Qy   1971 CAGAGAAGGTTGATTGCAGCGGCTTCACCACCATGCTGTTGAAACAGGAACCCCAAAAGGCAAC 2030
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   80 CAGAGAAGGTTGATGCAGCGGCTTCACCACCATGCTGTTGAAACAGGAACCCCAAAAGGCAAC 21

Qy   2031 ATTTGGCCTGTTTAGGTACC 2050
          ||||||||||||||||||||
Db   20 ATTTGGCCTGTTTAGGTACC 1

RESULT 10
ID   ABV36125/c
AC   ABV36125;
XX
AC   ABV36125;
XX
DT   16-SEP-2002 (first entry)
XX
DE   Human prostate expression marker cDNA 36116.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacogenomic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 200US-183319P.
PR   16-MAR-2000; 200US-189862P.
PR   25-MAY-2000; 200US-207454P.
PR   09-JUN-2000; 200US-211314P.
PR   18-JUL-2000; 200US-219007P.
PR   13-DEC-2000; 200US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R, Endege WO, Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS   Claim 1; Page 7483; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising

CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement; (I) is useful with prostate cancer;
CC   (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer in a patient;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
SQ   Sequence 439 BP; 113 A; 104 C; 110 G; 112 T; 0 other;

Query Match   9.3%; Score 288.8; DB 23; Length 439;
Best Local Similarity   91.4%; Pred. No. 6.8e-62;
Matches 338;   Conservative   0; Mismatches   7; Indels 25; Gaps   2;

Qy   1687 GATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAAGATGCACTTCGG 1746
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   439 GATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAAGATGCACTTCGG 380

Qy   1747 TCTTTTGTCAGATCCTCCGAGTGGCTTTCCCCAAGTGCTACCTATATAATAATCACCAGG 1806
          |||||||||||||||| |||||||||||||||||||||||||||||
Db   379 TCTTTTGTCAG--TCCTCCGGGTGCTTTCCCCAAGTGCTACCTATATAATAATCACCAGG 322

Qy   1807 CTAAGGACTGTATCGACTCCTTTGTTACTCACTGTTCGGCCATTCGTGATTATTCTC 1866
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 CTAAGGACTGTATCGACTCCTTTGTTACTCACTGTTCGGCCATTCGTGATTATTC 262

Qy   1867 AGATCCATGGACATAACCAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAAT 1926
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 AGATCCATGGACATAACCAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAAT 202

Qy   1927 TTGCCACCTTGCAGGATGAGTTTATGACATTTATTTTAATTAGGCAGAGAAGGTTGATGC 1986
          ||||||||||||||||||| |||||||||||
Db   201 TTGCCACCTTGCAGGATG----------------AGGCAGAGAAGGTTGATGC 165

Qy   1987 AGCGCTTCACCACCATGCTGTTGAAACAGGAACCCCAAAAGGCAACACATTTGGCCTGTTTAGG 2046
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   164 AGCGCTTCACCACCATGCTGTTGAAACAGGAACCCCAAAAGGCAACACATTTGGCCTGTTTAGG 105

Qy   2047 TACCTGGGTC 2056
          ||||||||||
Db   104 TACCTCGGCC 95

RESULT 11
ID   ABV45182/c
AC   ABV45182;
XX
AC   ABV45182;
XX
DT   16-SEP-2002 (first entry)
XX
DE   Human prostate expression marker cDNA 45173.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacogenomic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 200US-183319P.
PR   16-MAR-2000; 200US-189862P.

PR   25-MAY-2000; 2000US-2074454P.
PR   09-JUN-2000; 2000US-2113114P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R, Endege WO, Monahan JE;
XX
XX
DR   WPI; 2001-662795/76.
XX
XX
SQ   Sequence 439 BP; 113 A; 104 C; 110 G; 112 T; 0 other;

PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer -

Claim 1; Page 8947; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising
a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
specification or its complement. (I) is useful for:
(a) assessing whether a patient is afflicted with prostate cancer;
(b) monitoring the progression of prostate cancer in a patient;
(c) assessing the efficacy of a test compound to inhibit prostate
cancer in a patient;
(d) assessing the efficacy of a therapy for inhibiting prostate cancer
in a patient;
(e) selecting a composition for inhibiting prostate cancer in a patient;
(f) assessing the carcinogenic potential of a compound;
(g) assessing whether prostate cancer has metastasized in a patient;
(h) assessing the aggressiveness or indolence of prostate cancer in a
patient; and
(I) is also useful as a pharmacodynamic or pharmacogenomic marker.

Query Match   9.3%;   Score 288.8;   DB 23;   Length 439;
Best Local Similarity   91.4%;   Pred. No. 6.8e-62;
Matches 338;   Conservative   0;   Mismatches   7;   Indels   25;   Gaps   2;

Qy   1687 GATAACAAAAAGGTTCTTTGGAACTCATCTCATGCAAGACACTAGGTGAAGATGCACTTCGG 1746
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   439 GATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACACATGGTGAAAGATGCACTTCGG 380

Qy   1747 TCTTTTTGTCAGATCCTCCGAGTGCTTTCCCCAAGTGCTACCTTATATAATAATCACCAGG 1806
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   379 TCTTTTTGTCAG--TCCTCCGGGTGCTTTCCCCAAGTGCTACCTTATATAATAATCACCAGG 322

Qy   1807 CTAAGGACTGTATCGACTCCTTTGTTCGGCCATTCTGTAGTCTTATTC 1866
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   321 CTAAGGACTGTATCGACTCCTTTGTTACTCACTGTGTTCGGCCATTCTGTAGTCTTATTC 262

Qy   1867 AGATCCATGGACATAACAGGGGCTCGACGAGAGAGGAGAGAAGGTTGAGGAAT 1926
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 AGATCCATGGACATAACAGGGGCTCGACGAGAGATAAGCTTGGTCATATTCTTGAGGAAT 202

Qy   1927 TTGGCCACCTTGCAGGATGAGTTTATGACATTTTATTTRAATAGGGCAGGAAGGTTGAGTGC 1986
          ||||||||||||||||||||||||||||||| |||||||||||||
Db   201 TTGCCACCTTGCAGGATG--------------------AGGCAGAGAAGGTTGAGTGC 165

SQ   Sequence 439 BP; 113 A; 104 C; 110 G; 112 T; 0 other;

RESULT 12
ID   ABQ51660/c
AC   ABQ51660;

Qy   1987 AGCGCGTTCACACCATGCTGTTGTTGAAACAGGAACCCCAAAGGCAACATTTGGCCTGTTGTTAGG 2046
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   164 AGCGCGTTCACACCATGCTGTTGAAACAGGAACCCCAAAGGCAACATTTGGCCTGTTTAGG 105

Qy   2047 TACCTGGGTC 2056
          ||||||||||
Db   104 TACCTGGCC 95

ID   ABQ51660 standard; DNA; 799 BP.

XX
DT   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 38251.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200021B532-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.

PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -

Claim 12; 56pp + Sequence listing; 56pp; German.

This invention describes a novel method for determining the degree of
methylation of a particular cytosine in a motif 5'-CpG-3', present in a
genomic sample of DNA. The sample is treated chemically to convert
cytosine (C) but not methylated C, to uracil, then part of the genomic
DNA that contains the target C is amplified to form a labeled amplicon.
The amplicon is hybridised to two classes, each with at least one
member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
and the degree of hybridisation to both classes is determined from the
label on the amplicon. From the ratio of labels hybridised to the two
classes of oligomers, the degree of methylation is calculated. The method
is used: (i) for diagnosis and/or prognosis of side effects of
therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
of the central nervous, cardiovascular, gastrointestinal and respiratory
systems etc., particularly by detecting mutations or single nucleotide
polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
types and for investigating cell differentiation. The method allows the
methylation status of many C residues to be determined simultaneously.
ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
method for determining the degree of cytosine methylation described in
the disclosure of the invention.

WPI; 2002-371829/40.

SQ   Sequence 799 BP; 122 A; 95 C; 262 G; 320 T; 0 other;

Query Match   8.9%;   Score 274.2;   DB 24;   Length 799;
Best Local Similarity   73.9%;   Pred. No. 4.1e-58;
Matches 348;   Conservative   0;   Mismatches   123;   Indels   0;   Gaps   0;

Qy   86 CTGCTCGTTCGTTCGTTGGCTTCGCGAACCACTAAGGTCTACGGCAAACCTCCACGGTTTCCTTCCGCC 145
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   799 CTACTCGTTCGTTCGTACTGCGACCAACCACTAAAAATCTACGCAAAACCTCCACGGTTTCCTTCCGCC 740

Qy   146 TTCGGCTCACCTTTCTTAAGAAAATTCCCAGAGGGCAGGCCAGACGGGGGGGGGCTCTGAGAC 205
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   739 TTCGCGTCACCTTTCTTAAAAAAATTCCCAAAAAACAACGCCAAACGAAAACTCTAAAAC 680

Qy   206 TCCGGGCTCGCCGCTCTTCCGGGAACCGGCCCAGGACTACCAGAGACTGCCGAGAGGGTGAAA 265
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   679 TCCGAACTCCGCCCTCTTTCCCAAAAACCGCCCACTACCCAAAAACTCCGACAAAAAAATAAAA 620

Qy   266 AAGATAACTTCCGGTCTCGGGATCCGTCTCTRAATCTCGGGAGGAGAAGAGGCGCGCCA 325

```
Db   619 AAAAATAAACTTCCGATCTCGGATCGTCTAATCTCGCGAAAAAAAAGGACACGCA 560
         |||||  |||||||||||||| ||||||||| |||||||||  ||||
Qy   326 TCGGCCCAACGGACGGAGGCGGTGGCGGAGGGGGGTTGTGGCGGGGAGCGCGAAGTCCCGG 385
Db   559 TCGACCGAACGAAACAGATAACGAAAAAAATATAACCGAAAAAACGCGAAATCCCCGA 500
         ||||    |||||  |||||                 |||||||||   |||||||||
Qy   386 GAGTAAGGAGAGAGGGGGGGGGGGGGGGGGGGGGTCGGCCCATAGCCATGCCGTCGCACGCTGCGG 445
Db   499 AAATAAAAAAAAAAAAAAAACGAAATCGGCCGCCATTACGTACACGCTACCG 440
         |||||||||||||||||||   |||||                 ||||||||
Qy   446 GTCGGGGCTGGGGCTGAGAGGGGGGGGGGGGGGGGGGGGGGGGGCGCGGGCTCTGAGAC 505
Db   439 ATCGAACTAAACTAAAAAAAAAAACGACGACGACGACGTCGTTATTTCCGT 380
         |||||||||||                 |||||   ||||||||||||||
Qy   506 GGTCCGGACAGTGCGTGGGGGGGGGGGGGGGGGGGGGGGGAGTAGGCATAA 556
Db   379 AATCCGAACAATACGTAACGACGCGAATAACCACGAAAAAATAAATAAA 329
Qy   379 AATCCGAACAATACGTAACGACGCGAATAACCACGAAAAAATAAATAAAA 329
```

RESULT 13
ABQ51661

ID   ABQ51661 standard; DNA; 799 BP.
XX
AC   ABQ51661;
XX
DT   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 38252.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridisation of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of hybridisation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer,
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 799 BP; 320 A; 262 C; 95 G; 122 T; 0 other;

Query Match   73.9%; Score 274.2; DB 24; Length 799;
Best Local Similarity 8.9%; Pred. No. 4.1e-58;
Matches 348; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
Qy    86 CTGCTTCGTTCGTTCGTTCGTTCGCGCAACCACTRAGGTCTACGCAAACCTCCACGGTTTCCTTCGCC 145
         |||||||||||||||  ||||| || |||||||| |||||||||||||| ||||||||||||||
Db     1 CTACTTCGTTCGCTCGCTAACACTAAAAATCTAAGAG...CTAAGACACAACTTCCTTCGCC 60
Qy   146 TTCGGGTCACCTTTCTTAAGAAATTCCCGAGAGGGGCAGCAGCGCGGCAGCGGCGCTCTGAGAC 205
Db    61 TTCGGGTCACCTTTCTTAAAATTCCCAAAAAACAACAGCGAAACTCCACGAATTCCTTCGCC 120
Qy   206 TCCCGGCGCTCCGCCTCCTCTTTCCGGGAACACGGCCCACTACCCAGGACTCCGACAGAGGGGTGAAA 265
Db   121 TCCGAACTCCGCCTCCTTTCCGAAAAACGGCCCACTACCCAGAAAACCTCCGACAAAATAAAA 180
Qy   266 AAAGATTAACTTCCGGTCGTTCGGCGAATCGTCTCCTAATCTCGCGAGAAGAGAGAGGGGGCCGCA 325
Db   181 AAAAATAAACTTCCGATCTCGCGATCGTCTAATCTCGCGAAAAAAACGACACGCCA 240
Qy   326 TCGGCCCAACGGACGGAGGCGGTGGCGGAGGGGGTTGTGGCGGGGAGCGCGAAGTCCCGG 385
Db   241 TCGACCGAACGAAACAGATAACGAAAAAAATATAACCGAAAAAACGCGAAATCCCCGA 300
Qy   386 GAGTAAGGAGAGAGGGGGGGGGGGGGGGGTCGGCCCATAGCCATGCCGTCGCACGCTGCGG 445
Db   301 AAATAAAAAAAAAAAAAAACGAAATCGGCCGCCATTACGTACACGCTACCG 360
Qy   446 GTCGGGGCTGGGGCTGAGAGGGGGGGGGGGGGGGGGCGCGGGCTCTGAGAC 505
Db   361 ATCGAACTAAACTAAAAAAAAAAACGACGACGACGACGTCGTTATTTCCGT 420
Qy   506 GGTCCGGACAGTGCGTGGGGGGGGGGGGGGGGGGGAGTAGGCATAA 556
Db   421 AATCCGAACAATACGTAACGACGCGAATAACCACGAAAAAATAAATAAA 471
```

RESULT 14
ABV06162/c

ID   ABV06162 standard; cDNA; 303 BP.
XX
AC   ABV06162;
XX
DT   13-SEP-2002 (first entry)
XX
DE   Human prostate expression marker cDNA 6153.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer
XX
PS Claim 1; Page 1018; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 303 BP; 81 A; 72 C; 77 G; 73 T; 0 other;

Query Match   7.4%; Score 230.2; DB 23; Length 303;
Best Local Similarity 89.7%; Pred. No.2.5e-47;
Matches 279; Conservative   0; Mismatches   8; Indels 24; Gaps   2;

```
Qy 1746 GTCTTTTGTCAGATCCTCCGAGTGCTTACCTATATATAATAATCACCAG 1805
         ||||||||||||| |||||||||||  |||||||||||||||||||||
Db  303 GTCTTTTGTCAG-TCCTCCGGTGCTTCCCCCAGTGCTACCTATATAATAATCACCAG 245

Qy 1806 GCTAAGGACTGTATCGACTCCCTTTGTTACTCACTGTGTTCGCGCCATTCTGTAGTCTTATT 1865
         ||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db  244 GCTAAGGACTGTATCGACTCCCTTTGTTACTCACTGTGTTCGGCGCCATTCTGTAGTCTTATT 185

Qy 1866 CAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAA 1925
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  184 CAGATCCATGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAGGAA 125

Qy 1926 TTTGCCACCTTGCAGGATGAGTTTATGACATTTTATTTAATAGGCAGGAAGGTTGATG 1985
         |||||||||||||||||||                     |||||||||||||||||
Db  124 TTTGCCACCTTGCAGGATG------------------AGGCAGGAAGGTTGATG 88

Qy 1986 CAGCGCTTCACACACCTGTGTTGAAACAGGAAAACCCAAAGGCAAACATTTGGGCTTGTTTAG 2045
         || ||||||||| ||| |||||||||||||| | |||||||||| ||||||||||| |||||
Db   87 CACGCGTTCACACCATGCTGTTGAAACAGGAACCCCAAAGGCAACACATTTGGGCCTGTTTAG 28

Qy 2046 GTACCTGGGTC 2056
         ||||| |  |
Db   27 GTACCTCCGCC 17
```

XX
PN WO20017104Z-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-1916379P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DP P-PSDB; ABB63040.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 15911; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC sequences (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2515 BP; 616 A; 715 C; 657 G; 527 T; 0 other;

Query Match   7.0%; Score 217; DB 23; Length 2515;
Best Local Similarity 47.8%; Pred. No.1.4e-43;
Matches 1031; Conservative   0; Mismatches 1035; Indels 93; Gaps 10;

```
Qy 639  CTGGGTGGACATTACCCAAGATTTTGAAGAAGCTTGTCAGGAAGAATTAAAGTTGGGAGAACT 698
        |||||||||| ||||||| |  |||||  || ||  ||||| |  ||||  ||| |||||||
Db 350  CTGGGTGGACGTAACCAAGGAGAGTTCCACGATGCCTGCGGGAGCTGCAGCCGGGGGAACT 409

Qy 699  ACTTCATGATAAGCCTATTTGGTCTTTTTGAAGCCATGTCTGCTATTGAAATGATGGATCC 758
        |   ||||||  |||| ||||||||||||||||||| ||| |||    ||| ||||||||
Db 410  GGCCCAGGATATGCGTTCGGTCTTTTTGAGGCCATGCAGCCATCGAGATCATGGACCC 469

Qy 759  CAAGATGGATGCTGGCCATGATGGAAAACCAAGTTAATCGAAAGTTCTCAATTTTGAACA 818
        |||||||||  ||GGC   |  ||    ||||     ||||||  || |||    |||
Db 470  CAAGATGGACGTGGGCGCCCATCGGAGGTCGACAAGCAGGATCTTCCGCGCGCCCCTCTTTCGAGGC 529

Qy 819  AGCTATCAAGGATGGGCACTATTAAAAATTAAAGATTCACTTGCCTGAAACTGATAGGGAT 878
        ||| || || ||GGGC  ||| |   | ||  |||||||| |  ||| |||||||| | | |
Db 530  AGCCATTGCCACGGGCCCCATCAAAACTGACAGATCCACGCCTCCGAACTGAATGGCCAT 589

Qy 879  TATGGCATACATGTTTTTGCCTGTTTGATAAACGTGGTTTAGAAGGCCAATTCACTCGCACAGAC 938
        ||| | |  || ||||| ||||   ||   |||  ||    || || |||||||| ||| ||||
Db 590  CTATGAATGCTCTGTTTTTCCTGCCTGGGCGTCCCTGACGGGCAACTCCATGGACCAGGT 649

Qy 939  AGTATTTACGTGCCTTTACATTCAATAATCCAGACTTTATAGAAGATCCTGCTAGGAAGGC 998
        | |||| ||| |||| | |    | |  || |  || ||  ||||   |||| |||||  |
Db 650  TCTATTCACCTGCCTCTCACCTGCACGCTCCCGCACAGATCAAGGATAAGGGCGTGCCGT 709

Qy 999  TTTTGCTCTGGGAATCTTTGAAAATCTGTGACATGTCAAGGAAAAAGTAAATAAAAGCTGC 1058
        ||| ||| ||  ||||| | |||   || | ||  || || ||||| |  | | | | ||
Db 710  CTTCTCGCACTGCCGTGCGCCAATCTCATCGTGGTCATCAAGAAAATATTATTGCCGTGGCAGC 769

Qy 1059 TGTTTTTGAAGGAGGAGCAGAGATTTCAGTCAATGACTTATGGATTTAAAATGGCTAACAGTGT 1118
        ||| | |||||||||| | || | |||| | ||  || |  ||  |||||   ||  ||| ||
Db 770  TGTCAACGAGGAGGAGGAGGACTTCCAGCTATCACGGAAAATTCTGCTCTTCTCGGCGTGCGCAGAGAA 829
```

RESULT 15
ID   ABL07143 standard; cDNA; 2515 BP.
AC   ABL07143;
DT   26-MAR-2002 (first entry)
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 15911.
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
OS   Drosophila melanogaster.

```
Qy  1119 GACAGATC---TTCGAGTTACAGGCATGCTAAAAGATGTGGAGATGACATGCAAAGAAG 1175
         |||||||    |||||||||||||||||  |||||||||||||||||||||  |
Db   830 GGCACAGCCAGCCACTGTTTATAGCTCGTCTAAAAGATGTGGAGATGAGCTCATACG--- 886

Qy  1176 AGTAAAGAGTACTCGACAAGGAGAAGAAAGATCCAGAAGTTGAACTAGAACA 1235
                        |||||  |||
Db   887 ---------------CAAGTGCAAAAAGCTGACTTCCACGGAGGACTGGATGG 924

Db   925 CCGTGGTTCACCGGCTGCGTT--TCATGCGGCCATCTTTTCCAGGTGATCTACCACGTGGA 982
Qy  1296 TATAGCCTTTACTAAGAAAAGAGACCAGTGGTTGCAGAAGCTCAAAAAATGATTGGTTCA 1355

Qy  1236 CCAACAAATGTTTAGCAGTATTCAGCGAGGTGAAAATTTACTCGTGTTGACAGTGCT 1295
Db   983 GCAAATGGCCAGCAAATGATCACGTGGTGAGTGAGATATACAAAAATCCTGCTTGT 1042

Db  1043 GGGCTCCGAGATGCTTCCAGGAATAAAGGAGCACACTTTGGATCGGGGCACACCACAGAGA 1102
Qy  1356 AGCAGCAGATCTTCTTTCTCCCATTCATAATTCATTCGCATCCTAGGCCTTCCCAGGAA 1415

Qy  1416 TGATACTACAAAAGGAGGATCATCCAATTATCAATTATGATAGTGG... 1475
Db  1103 GGGGTCT-------GACGGCTCCCAATCCCATTGGGTTTCTCTCCGCGGATCCACGACCG 1153

Qy  1476 GCTCACTTCCACCTACCTTCCCTCGATATGCAAAAATAATTAAAAGGGAAGGAAAATGGTGAA 1535
Db  1154 CAGGCAACCGCCCCGGTTTCCGCGTAGCCATTAAGATCACGTCATCATCCAGTTATCA 1213

Qy  1536 CTATTTTGCAAGATTAATAGAGAATAAAAAACTGTCTGTGAGGTTGTGAATTAACAAA 1595
Db  1214 GTTTCTAGAGGAAATGATTTGCGCTTCAAAGTCACCAAGTACAAGGA 1273

Db  1274 TTACTATTCGGCGTGAACTCTTTATCGAGTGACTAGCAAAAAAGTCGGGCCAGTGCATCCT 1333
Qy  1596 TTTACATTGTATCCTGGGATTTTTTTCTGTGAATTTAGTGAACAGTC---ACCATGTGTTCT 1652

Db  1334 GTCCAGAAGCCTGCTGCAAACCCTGTTCAGCGCGCCAACATGCGTATGGCGCACGGAAAGCT 1393
Qy  1653 TTCAAGATTCTCTTGTTACAAACCACTTTCCTGGTGGGATAACAAAAAGGTCTTTGGAACTCA 1712

Qy  1773 TCCCCCAAGTGCTACCTATATAATAATCACCAGGCTAAGGACTGTATCGACTCCTTTGTT 1832
Db  1394 TCCCATGAAGCAGTTCCTGCCGCCACTCGGTTCAACTGCCACCGTATTGAA 1453

Db  1454 T--GCCAAGCATCCGGTGGCTCGCCGATCCCAAGGTCAGCGACGAACTTCTTC 1511
Qy  1833 ACTCACTGTGTT------CGGCCATTCTGTGTAGTCTTTATTCAGATCCATGGACATAAACAGG 1886

Qy  1887 GCTCGACAGAGATAAGCTTGGTCATATTCTTGAGGAATTGCCACCTTGCAGGATGAG 1946
Db  1512 CGTTACTGCATTAACATGAACACACTTCAGCGCAGTTCATTCGCATCTGCGGCTTCAAATCGC 1571

Qy  1947 TTTATGACATTTATTTTAATAGGCAGAGAGGTTGAATGCAGGCGCTTCACACCATGCTGT 2006
Db  1572 GCCAGACAGCGCGGGATAAGCGTCTGAATAGAGAAACTTCGACACCATCCAAGTGGAT 1631

Qy  2007 TGAAACAGGGAACCCAAAG-GCAACATTGGCCTGTTGAGTACCTGGGTCCTTTACCAT 2065
Db  1632 GCAGCGCGGATTGGACTCTATGATGAACAACTGGCCAACGAGCCATGGAGGGCAAC 1691

Db  1692 GAGCCCATGGCCACCGCTCTGAAGCACGCACCACTTCTCCACATTGGTTCGTTGTCAAC 1751
Qy  2066 AACCTTCGCATTATGATACAGTACCTTCTAAGTCTTGAATTGGAACTCTACAGTATG 2125

Db  1752 TGTTTTCCGTGCGAATGCTGATCTTCCTGATGTCCGGCTTTGGAACTATATGCTGTG 1811
Qy  2126 CACGAGTACTATTACATATATATTGGTATATCTCTGAATTCCTTTACGGATGTTGATGTCA 2185

Qy  2186 ACATTGAGTCGTGCCGATGGCTCTCAAATGGCAGAGAGGAAAAGGATAATGGAAGAGCGCAG 2245
Db  1812 CATGAGTTCCTCTACATCTTACTCGGTTATCCATACGAGCTTCCTCGGTCTCC 1871
```

```
Db  1872 GCGCTGACGCGCACGGAGAACATCCTCCTGGCCGCCAGGAGGAGTACGCGCCAAAGT 1931
Qy  2246 AAAGGCCGTAGTAG-----------------TAAAAACAAAG 2272
Db  1932 AAGACCCAGTCGGGCGGGGACGCGGGCGCTAAAAACCGGCAAGCCCAAGCCGAAAAG 1991

Qy  2273 AAAAAAAAGAAAAAGTTCGCCCATTGACCGGAGGAGGATCACAAATGAGCCAACATATCAGAAC 2332
Db  1992 AACAAAAAGACTCAAAAGCCCGTACCCGCGGAGAGATCGTCTTCTTACCACGCGGCTACTCAGC 2051

Qy  2393 CCGAAGTTTGAGCTTGAATAGTGAACAAGTTCGGTATGAAGAACAGTTGCTCATTCAAC 2452
Db  2052 ATGTGTCGCTGGAATGTTTAAAACCATGGTAGCATTGACGGCGAAATGTACGTAA 2111

Db  2112 CCGTGTCCAAGTTCGACAACGGAGGAGATCCGCTACAACCACAGAGAA 2171
Qy  2453 AGTGTGATGACCCCGCCGCCCAGTGCACTACTTACAGTGTCAAGGAAAATGTCTGACTTCAAT 2512

Qy  2513 AAAATATAGCCCTCCTCCCTCAGTTCCCTGAAACTGTATGTGGGCAGCTAGTAAGCACTTTCAA 2572
Db  2172 ACACTCACCACCAGTCCCTCCGCCAGTTTCGTACGCACAGAGTTCAAAAAGGTCCCGGGAGCACATG 2231

Db  2232 A--------TGCGACATAGCCTGCGAGGATCTTATATACCTTATGCGGCAAGCACTTTGAC 2282
Qy  2573 CAGGCAAAAATGATATTGGAAAATATTCCTAACCCGGACCATGAGGGTTAATAGAATTTTA 2632

Qy  2633 AAGGTTGCAAAACCCAACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAGGAA 2692
Db  2283 CAGGCCCGCAATGTTCTTCGAAAGCATTCAAAAACCCAGATCAGGAGGAGGATGCTGGACTCCTG 2342

Db  2343 CAAATCGGCACGGACTAACTTTGTTGGTGATGAATGGTGCTGCGGCGTCACCAGAAGGAA 2402
Qy  2693 TCTAAAGTTCCTCCTGAATTTGATTTCTCTGCTCAATAAATATTTTTCCTGTTGTCTGAAAACT 2751

Db  2403 GTCAAGGGGCAGCCAGCGAGTTCGACTTCTCCCAAACACAGCTACTTCCCCATACATCAAGCT 2461
```

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.6

Run On: November 24, 2003, 21:46:30 ; Search time 11138 Seconds
             (without alignments)
             11371.539 Million cell updates/sec

Title: US-10-001-857-42
Perfect score: 3096
Sequence: 1 ttcctcacgaaactcccagg...........atttcttgacaaaaaaaaa 3096

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
                 1: gb_ba:*
                 2: gb_htg:*
                 3: gb_in:*
                 4: gb_om:*
                 5: gb_ov:*
                 6: gb_pat:*
                 7: gb_ph:*
                 8: gb_pl:*
                 9: gb_pr:*
                 10: gb_ro:*
                 11: gb_sts:*
                 12: gb_sy:*
                 13: gb_un:*
                 14: gb_vi:*
                 15: gb_ba:*
                 16: gb_fun:*
                 17: em_hum:*
                 18: em_in:*
                 19: em_mu:*
                 20: em_om:*
                 21: em_or:*
                 22: em_ov:*
                 23: em_pat:*
                 24: em_ph:*
                 25: em_pl:*
                 26: em_ro:*
                 27: em_sts:*
                 28: em_un:*
                 29: em_vi:*
                 30: em_vrt_hum:*
                 31: em_htg_inv:*
                 32: em_htg_other:*
                 33: em_htg_mus:*
                 34: em_htg_pln:*
                 35: em_htg_rod:*
                 36: em_htg_mam:*
                 37: em_btg_vrt:*
                 38: em_sy:*
                 39: em_htg_hum:*
                 40: em_htg_mus:*
                 41: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	3096	100.0	3096	6	AX523454 AX523454 Sequence
2	2417.2	78.1	2488	6	AX537452 AX537452 Sequence
3	1838.4	59.4	2552	10	AY102701 AY102701 Mus muscu
4	1799.2	58.1	2543	10	AF272892 AF272892 Rattus no
5	1766.6	57.1	1839	9	AK025266 AK025266 Homo sapi
6	1264.8	40.9	2467	5	BC053286 BC053286 Danio rer
7	988	31.9	997	9	AK026296 AK026296 Homo sapi
8	885.4	28.6	1796	9	AK056059 AK056059 Homo sapi
9	717.4	23.2	719	2	AX664135 AX664135 Sequence
10	664.2	21.5	147876	2	AC102010 AC102010 Mus muscu
11	551.4	17.8	181864	9	AL353743 AL353743 Human DNA
12	398.1	12.9	160115	9	AL161447 AL161447 Human DNA
13	321.4	10.4	446	9	HSA334821 HSA334821 Homo sapi
14	299	9.7	743	9	HSA334413 HSA334413 Homo sapi
c 15	236.4	7.6	642	9	HSA342328 HSA342328 Homo sapi
c 16	236.4	7.6	742	9	HSA331811 HSA331811 Homo sapi
c 17	233.2	7.5	640	9	HSA342329 HSA342329 Homo sapi
c 18	233.2	7.5	652	9	HSA342327 HSA342327 Homo sapi
19	217	7.0	2537	3	AY060948 AY060948 Drosophil
20	207.6	6.7	2565	3	AY089452 AY089452 Drosophil
21	207	6.7	320902	11	AL161456 AL161456 Homo sapi
c 22	182	5.9	489	11	G63197 G63197 SHGC-140830
c 23	177.8	5.7	34935	9	AL161453 AL161453 Human DNA
c 24	177.6	5.7	111554	2	AC143812 AC143812 Macaca mu
c 25	155.8	5.7	113836	10	AL669837 AL669837 Mouse DNA
c 26	153.2	4.9	253611	2	AC111473 AC111473 Rattus no
c 27	153.2	4.9	271861	2	AC108541 AC108541 Rattus no
c 28	153.2	4.9	281804	2	AC134869 AC134869 Mus muscu
c 29	140.4	4.5	151	2	AX523453 AX523453 Sequence
c 30	134.4	4.3	204131	10	AC128898 AC128898 Rattus no
c 31	130.4	4.2	3497	2	AY102702 AY102702 Mus muscu
32	130.4	4.2	269131	2	AC135469 AC135469 Mus muscu
c 33	90.6	2.9	16016	2	AC020508 AC020508 Drosophil
34	90.6	2.9	60135	2	AC091200 AC091200 Drosophil
c 35	90.6	2.9	164443	3	AC012167 AC012167 Drosophil
36	90.6	2.9	298827	3	AE003463 AE003463 Drosophil
c 37	87.8	2.8	815	6	AX523523 AX523523 Sequence
c 38	69.4	2.2	7218	6	I66494 I66494 Sequence 14
c 39	69.4	2.2	2000	6	AX655393 AX655393 Sequence
c 40	66.6	2.2	110737	2	AC011105 AC011105 Homo sapi
c 41	66.2	2.1	159980	2	AP005743 AP005743 Oryza sat
c 42	65.6	2.1	68328	2	AC137705 AC137705 Homo sapi
c 43	64.6	2.1	832	11	BV026280 BV026280 S212P6647
c 44	64.4	2.1	69373	2	AC101456 AC101456 Mus muscu
c 45	64	2.1	67769	2	AC068336 AC068336 Homo sapi

ALIGNMENTS

RESULT 1
AX523454
LOCUS   AX523454   3096 bp   DNA   linear   PAT 24-OCT-2002
DEFINITION   Sequence 42 from Patent WO02064788.
ACCESSION   AX523454
VERSION   AX523454.1 GI:24412250
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Macina,R.A., Recipon,H., Chen,S.Y., Sun,Y. and Liu,C.
  TITLE   Compositions and methods relating to lung specific genes and
              proteins

```
JOURNAL      Patent: WO 0206478-A 42 22-AUG-2002;
             Diadexus, Inc. (US)
FEATURES             Location/Qualifiers
     source      1..3096
                 /organism="Homo sapiens"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:9606"
BASE COUNT    914 a    606 c    749 g    827 t
ORIGIN

Query Match          100.0%;  Score 3096;  DB 6;  Length 3096;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 3096;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
Qy      1 TTCCTCACGAAAACTCCCAGGCGCCTGTATAGGAAAACATAAATCCGTTGTCAGGCAGAGTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTCCTCACGAAAACTCCCAGGCGCCTGTATAGGAAAACATAAATCCGTTGTCAGGCAGAGTA 60

Qy     61 GCACGGCTGTTGCTCTCCGGAGCTTGGGCTCTTGGCTGCTCGCAACCACTAAGGTCTACG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCACGGCTGTTGCTCTCCGGAGCTTGGGCTCTTGGCTGCTCGCAACCACTAAGGTCTACG 120

Qy    121 CAAAACCTCCACGGTTTCCTTCCGCCCTTCGGCCTCACCTTTCTTAAGGAAATTCCCAGAGGGCA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAAAACCTCCACGGTTTCCTTCCGCCCTTCGGCCTCACCTTTCTTAAGGAAATTCCCAGAGGGCA 180

Qy    181 GCGCAGAGACGGGGGGCGGGGGGCTCTGAGACTCCCGGCGCTCCGCGGGAAACCGCCCACTA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCGCAGAGACGGGGGGCGGGGGGCTCTGAGACTCCCGGCGCTCCGCGGGAAACCGCCCACTA 240

Qy    241 CCCAGGACTCCGACAGAGGGTGAAAAAAGAGATAACTTCGGGTCTCGCGGATCGTGTCTTCTAATC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCCAGGACTCCGACAGAGGGTGAAAAAAGAGATAACTTCGGGTCTCGCGGATCGTGTCTTCTAATC 300

Qy    301 TCGCGGAAGAGAAGAGGCGCGCGAAGTCCCGGGACTAAGGGAGAGAGGGCGGGGGTCCCGG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCGCGGAAGAGAAGAGGCGCGCGAAGTCCCGGGACTAAGGGAGAGAGGGCGGGGGTCCCGG 360

Qy    361 TGGCCGGGGGAGCGGGGAAGTCCCGGGGAGTAAGGGAGAGGGGCGGGGGTCGGCGGTCCCGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGCCGGGGGAGCGGGGAAGTCCCGGGGAGTAAGGGAGAGGGGCGGGGGTCGGCGGTCCCGG 420

Qy    421 GCATACGCATGGCGTGCACGCGCGCGGTCGGGCGTCGGGCGTCGAGGGGGAGGGGCGGGGGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GCATACGCATGGCGTGCACGCGCGCGGTCGGGCGTCGGGCGTCGAGGGGGAGGGGCGGGGGG 480

Qy    481 GCCGAGGGGGCGGGGTCGTTGTTTCGGTGGTCCGGACAGTGCGGTCGCGGTGACCACG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCCGAGGGGGCGGGGTCGTTGTTTCGGTGGTCCGGACAGTGCGGTCGCGGTGACCACG 540

Qy    541 GGAGAAGTAGGGCATAAATGGTTAATTGGGATTCTTCGTTGTAGGATCAGGGATGGGAG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GGAGAAGTAGGGCATAAATGGTTAATTGGGATTCTTCGTTGTAGGATCAGGGATGGGAG 600

Qy    601 CTCAGTATGCCAGAAAAAATGGAGAGAAAAGCAAATACAAAACTCGGGTGGGACTTACCCAAGAT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTCAGTATGCCAGAAAAAATGGAGAGAAAAGCAAATACAAAACTCGGGTGGGACTTACCCAAGAT 660

Qy    661 TTTGAAGAAGCTTGTTCGGAGAATTAAAGTTGGGAGAACTACTTCAATGATAAGCTATTTGGT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTTGAAGAAGCTTGTTCGGAGAATTAAAGTTGGGAGAACTACTTCAATGATAAGCTATTTGGT 720

Qy    721 CTTTTTGAAGCCATGTCTGCTATTGAAATGATCGGATCCCAAGATGGATGGCTGGCATGATT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CTTTTTGAAGCCATGTCTGCTATTGAAATGATCGGATCCCAAGATGGATGGCTGGCATGATT 780

Qy    781 GGAAACCAAGTTAATCGAAAAAGTTCTCAAATTTGAACAAGCTATCAAGGACACTATT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGAAACCAAGTTAATCGAAAAAGTTCTCAAATTTGAACAAGCTATCAAGGACACTATT 840

Qy    841 AAAAATTAAACAGATCTCACCTTGCCTGAACTGATAGGGATTATGGGATAACATGTTTTTGCTGT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AAAAATTAAACAGATCTCACCTTGCCTGAACTGATAGGGATTATGGGATAACATGTTTTTGCTGT 900
```

```
Qy    901 ATCGTGACATTGCAAGGAGAAAAAGTAAAAAGTAAATAAATAAAAGCTGCTGTTTTTTGAAAGATTTT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ATCGTGACATTGCAAGGAGAAAAAGTAAAAAGTAAATAAATAAAAGCTGCTGTTTTTTGAAAGATTTT 960

Qy    961 TTGATAAACGTGGTTAGAAGGGCCATTCACTGGCACAGACAGACAGTAGTTACGTGCCTTTACATT 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTGATAAACGTGGTTAGAAGGGCCATTCACTGGCACAGACAGACAGTAGTTACGTGCCTTTACATT 1020

Qy   1021 CAGTCAATGACTTATGGATTAAAATGGGCTAACAGTGTGACAGATCTTCGAGTTACAGGC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CAGTCAATGACTTATGGATTAAAATGGGCTAACAGTGTGACAGATCTTCGAGTTACAGGC 1080

Qy   1081 ATGCTAAAAGATGTGGACAGATGACATGCAAAGAAGAGTAAAGAGTACTCGAACAA 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ATGCTAAAAGATGTGGACAGATGACATGCAAAGAAGAGTAAAGAGTACTCGAACAA 1140

Qy   1201 GGAGAGAAAAGAGATCCAGAAGTTGAACTAGAACACAACAAATGTTTTAGCAGTATTCAGC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GGAGAGAAAAGAGATCCAGAAGTTGAACTAGAACACAACAAATGTTTTAGCAGTATTCAGC 1260

Qy   1261 AGAGTGAAAATTTACTCGTGTGTTACTGACAGTGCTTATAGCCTTTACTRAAGAAAGAGACC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 AGAGTGAAAATTTACTCGTGTGTTACTGACAGTGCTTATAGCCTTTACTRAAGAAAGAGACC 1320

Qy   1321 AGTTGCTGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATT 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AGTTGCTGTTGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTCTGCCATT 1380

Qy   1381 CATATTCATTGCATCATGGCATCCAGGCCAGAATGATAATTGATTACTACAAAAAGGACATCATCCA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 CATATTCATTGCATCATGGCATCCAGGCCAGAATGATAATTGATTACTACAAAAAGGACATCATCCA 1440

Qy   1441 ATTATGATGGGTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACCTTCCCTCGA 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 ATTATGATGGGTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACCTTCCCTCGA 1500

Qy   1501 TATGCAAAAATAATTAAAAGGAAGAAAATGGTGAACTATTTTGCAAGATTAATAGATAGA 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 TATGCAAAAATAATTAAAAGGAAGAAAATGGTGAACTATTTTGCAAGATTAATAGATAGA 1560

Qy   1561 ATAAAAACTGTCTGTGAGGTTGTGAGATTTAACAAAATTACATTGTATCCTGGATTTTTTTC 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ATAAAAACTGTCTGTGAGGTTGTGAGATTTAACAAAATTACATTGTATCCTGGATTTTTTTC 1620

Qy   1621 TGTGAATTAGTGAACAGTCAACATGTTTCTTTCTTGTTACAAAACCACTTTC 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TGTGAATTAGTGAACAGTCAACATGTTTCTTTCTTGTTACAAAACCACTTTC 1680

Qy   1681 CTGGTGGATAACAAAAAGTTCTTTGGAACTCATCTCATTGCAAGAACGAAGAGATGCA 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CTGGTGGATAACAAAAAGTTCTTTGGAACTCATCTCATTGCAAGAACGAAGAGATGCA 1740

Qy   1741 CTTCGGTCTTTTGTCAGATCCTCGAGTGCTTTCCCCAGTGCTTACCCTRATATAATTAATC 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CTTCGGTCTTTTGTCAGATCCTCGAGTGCTTTCCCCAGTGCTTACCCTRATATAATTAATC 1800

Qy   1801 ACCAGGCTAAGGACTGTATCCGACTCCTTTGTTACTCACTGTTCGGGCCATTCGTAGTC 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ACCAGGCTAAGGACTGTATCCGACTCCTTTGTTACTCACTGTTCGGGCCATTCGTAGTC 1860

Qy   1861 TTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGTAAGCTTGGTCATATTCTTG 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TTATTCAGATCCATGGACATAACAGGGCTCGACAGAGAGTAAGCTTGGTCATATTCTTG 1920

Qy   1921 AGGAAATTTGCCACCTTGCAGGATGAGTTTATGACATTTATTTAATTAGGCAGAGAGAAGGT 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 AGGAAATTTGCCACCTTGCAGGATGAGTTTATGACATTTATTTAATTAGGCAGAGAGAAGGT 1980
```

```
RESULT 2
LOCUS       AX537452               2488 bp    DNA     linear   PAT 23-NOV-2002
DEFINITION  Sequence 23 from Patent WO02070709.
ACCESSION   AX537452
VERSION     AX537452.1  GI:25269238
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 2488)
  AUTHORS   Lal,P.G., Baughn,M.R., Yao,M.G., Walia,N.K., Elliot,V.S., Xu,Y.,
            Honchell,C.D., Yue,H., Ding,L., Gietzen,K.J., Ison,C.H., Lu,D.A.,
            Hafalia,A.J., Ghandi,A.R., Thangavelu,K., Sanjanwala,M.M.,
            Tang,Y.T., Ramkumar,J., Griffin,J.A., Swarnaker,A., Azimzai,Y.,
            Sapperstein,S.K., Burford,N., Lee,E.A., Lu,Y., Tran,U.K. and
            Marquis,J.P.
  TITLE     Molecules for disease detection and treatment
  JOURNAL   Patent: WO 02070709-A 23 12-SEP-2002;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..2488
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 7111920CB1"
BASE COUNT      768 a    461 c    567 g    692 t
ORIGIN

Query Match             78.1%;  Score 2417.2;  DB 6;  Length 2488;
Best Local Similarity   98.5%;
Matches 2470;  Conservative   0;  Mismatches  13;  Indels  25;  Gaps  2;
```

(DNA sequence alignment, Query (Qy) versus Database (Db); vertical bars indicate base matches)

```
Qy  474  GGCGCGGCGGCCGGAGGGCGGGGCGCGGGCGGGCGTCGGTTATTTCCGTGGCTCGGGACAGTTGCGTGCGCGCGGGGT  533
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  GGCGGCGGCCGGAGGGCGGGGCGCGGGCGGGCGTCGGTTATTTCCGTGGCTCGGGACAGTTGCGTGCGCGCGGGGT    60

Qy  534  GACCACGGGAGAGAAGTAGGGCATAATGGTTTATGAAAGCTTCGTAGATGGATTCAGG  593
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  GACCACGGGAGAAGTAGGCATAATGGTTTATGAAAGCTTCGTAGATGACGAATTCAGG  120

Qy  594  ATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAAACTGGTGGACATTAC  653
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  ATGGGAGCTCAGTATGCCAGAAAAAATGGAGAAAAAGCAATACAAACTGGGTGGACATTAC  180

Qy  654  CCAAGATTTTGAAGAAGCTTGTCGAGAATTAAAGTTGGGAGAACAACTACTTCATGATAAGCT  713
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  CCAAGATTTTGAAGAAGCTTGTCGAGAATTAAAGTTGGGAGAACAACTACTTCATGATAAGCT  240

Qy  714  ATTTGGTGCTTTTTTGAAGCCATGTGTCTGCTATTGAAATGATGATGGATCCCAAGATGGA  773
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  ATTTGGTGCTTTTTTGAAGCCATGTGTCTGCTATTGAAATGATGATGGATCCCAAGATGGA  300

Qy  774  CATGATTGGAAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAACAGCTATCAAGGATGG  833
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  CATGATTGGAAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGG  360

Qy  834  CACTATTAAAAATTAAAGATCTACCTTGCCTGAACTGATAGGGGATTATGGATACAGTTT  893
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  CACTATTAAAAATTAAAGATCTCACCTTGCCTGAACTGATAGGGATTATGGATACAGTTT  420

Qy  894  TTGCTGTTTGATAACGTGGTTAGAAGGGCCATTCACTGGCACACGACAGTATTAGTGCCT  953
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  TTGCTGTTTGATAACGTGGTTAGAAGGGCCATTCACTGGCACACGACAGTATTAGTGCCT  480

Qy  954  TTACATTCATAATCCAGACTTTATAGAAGATCCTGGTATGAAGGCTTTTTGCTCTGGGAAT  1013
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  TTACATTCATAATCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTTGCTCTGGGAAT  540
```

```
Qy  1981  TGATGCAGCGCGCTTCACCACCATGCTGTTGAAACAGGAACCCCAAGGCAACCATTGGGCCTG  2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981  TGATGCAGCGCGCTTCACCACCATGCTGTTGAAACAGGAACCCCAAGGCAACCATTGGGCCTG  2040

Qy  2041  TTTAGGTACCTGGGTCTCCTTTACCATAACCTTCGCATTATGATACAGTACCTTCTAAGTGG  2100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041  TTTAGGTACCTGGGTCTCCTTTACCATAACCTTCGCATTATGATACAGTACCTTCTAAGTGG  2100

Qy  2101  CTTTGAATTGGAACTCTACAGTATGCCACGAGTACTATTACATATAGTATCTCTCTCGA  2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101  CTTTGAATTGGAACTCTACAGTATGCCACGAGTACTATTACATATAGTATCTCTCTCGA  2160

Qy  2161  ATTCCTTTACGCATCGGTTGATGTCAACATTGAGTCGTCGTGCCGATTGGCTCCAAATGGCAG  2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161  ATTCCTTTACGCATCGGTTGATGTCAACATTGAGTCGTCGTGCCGATTGGCTCCAAATGGCAG  2220

Qy  2221  GGAAAGGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAAACAAAGAAAAAAA  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221  GGAAAGGATAATGGAAGAGCAGCAGAAAGGCCGTAGTAGTAAAACAAAGAAAAAAA  2280

Qy  2281  GAAAGTTCCCCATTGAGCGAGATCACAATGAGCCAAGCATATCAGAACATGTGTGC  2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281  GAAAGTTCGCCCATTGAGCCGAGATCACAATGAGCCAAGCATATCAGAACATGTGTGC  2340

Qy  2341  TGGAATGTTTAAAACCATGGTAGTCATTTGACATGGACAGCGGCGAAAGTACTGAACCGGAAGTT  2400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341  TGGAATGTTTAAAACCATGGTAGCATTTGACATGGACAGCGGCGAAAGTACTGAACCGGAAGTT  2400

Qy  2401  TGAGCTTGATTAGTAGTGAACAAGTTCGGTATGAACACACAGTGTGAT  2460
          |||||||||||||||||||||||||||||||||||||||||||||||
Db  2401  TGAGCTTGATTAGTAGTGAACAAGTTCGGTATGAACACACAGTGTGAT  2460

Qy  2461  GACCCCGCCCGCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAG  2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461  GACCCCGCCCAGTGCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAATATAG  2520

Qy  2521  CCCTCCTTCCTCCAGTCTCTCCTGAACTGTATGTGGCCAGCTAGTAGTGAAGCACTTTCAACAGGCAAA  2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521  CCCTCCTTCCTCCAGTCTCTCCTGAACTGTATGTGGCCAGCTAGTAGTGAAGCACTTTCAACAGGCAAA  2580

Qy  2581  AATGATATTGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTAAAGGTTGC  2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581  AATGATATTGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTAAAGGTTGC  2640

Qy  2641  CAAACCCAAACTTTGTGGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGT  2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641  CAAACCCAAACTTTGTGGGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGT  2700

Qy  2701  TCCTCCTGAATTTGAATTTCCTGCTCTCAATAAATATTTTTCCTGTTGTGAAACTTGTTTGAGA  2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2701  TCCTCCTGAATTTGAATTTCCTGCTCTCAATAAATATTTTCCTGTTGTGAAACTTGTTTGAGA  2760

Qy  2761  GAGACTGGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTCAGACCCAACTCTTAGAGGG  2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2761  GAGACTGGGGAGGTGGCCATAAAGGGGCAGAGTCTTCTTTCAGACCCAACTCTTAGAGGG  2820

Qy  2821  CACATCACCAGGCTCCACACATCACGGGAAGTGAGAATGGAGAATTCCTTGGGTAACAACTCATTA  2880
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2821  CACATCACCAGGCTCCACACATCACGGGAAGTGAGAATGGAGAATTCCTTGGGTAACAACTCATTA  2880

Qy  2881  TAAGGAAATACTTTTAGTTTGACAGCCTTTATATGACATGAATGAAAACTGCTGTTTTTAAAG  2940
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2881  TAAGGAAATACTTTTAGTTTGACAGCCTTTATATGACATGAATGAAAACTGCTGTTTTTAAAG  2940

Qy  2941  TGGTTTATTTATGTTCCATGGAAGAAACTGGTCTCTTATTGAATGCATTGACATTGAAGGTTATAT  3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2941  TGGTTTATTTATGTTCCATGGAAGAAACTGGTCTTTATTGAATGCATTGAATGAAGGTTATAT  3000

Qy  3001  GGTTTATTACAGAGATTTAAATCACAAAATCATTTTTTATGAATGATTGAGTGAAAAATAGTGT  3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  3001  GGTTTATTACAGAGATTTAAATCACAAAATCATTTTTTTATGAATGATTGAGTGAAAAATAGTGT  3060

Qy  3061  TTATAAAGGTTAATAAATTCTTGACAAAAAAAAAA  3096
          |||||||||||||||||||||||||||||||||||
Db  3061  TTATAAAGGTTAATAAATTCTTGACAAAAAAAAAA  3096
```

```
Qy    1014 CTTGAAAATCTGTTGACATTGCAAGGAGAAAAAGTAAATAAAGCTGCGTGTTTTTGAAGAGGA 1073
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1014 CTTGAAAATCTGTTGACATTGCAAGGAGAAAAAGTAAATAAAGCTGCGTGTTTTTGAAGAGGA 1073

Qy    1074 AGATTTTCAGTCAATGACTTATGGATTAAAAGGGCTAAACAGTGTGACAGATCTTCGAGT 1133
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGATTTTCAGTCAATGACTTATGGATTAAAAGGGCTAAACAGTGTGACAGATCTTCGAGT  660

Qy    1134 TACAGGCATGCTAAAAGATGTGGAGGATGACATGCAAAGAGAAGAGTAAAGTACTCGAAG 1193
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 TACAGGCATGCTAAAAGATGTGGAGGATGACATGCAAAGAGAAGAGTAAAGTACTCGAAG  720

Qy    1194 TCGACAAGGAGAAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGTTTTAGCAGT 1253
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TCGACAAGGAGAAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAATGTTTTAGCAGT  780

Qy    1254 ATTCAGCAGAGTGAAATTTTACTCGTGTGTTACTGAACAGTGCTTATAGCCTTTACTTAAGAA 1313
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ATTCAGCAGAGTGAAATTTTACTCGTGTGTTACTGACAGTGCTTATAGCCTTTACTTAAGAA  840

Qy    1314 AGAGACCAGTGCTGTTGCGAAGCTCAAAAATTGATGGTTCAAGCAGCAGAGATCTTCTTTC 1373
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 AGAGACCAGTGCTGTTGCGAAGCTCAAAAATTGATGGTTCAAGCAGCAGAGATCTTCTTTC  900

Qy    1374 TGCCATTCATAATTCATTCATTGCATCATGCACAGCCCAGAATGATACTACAAAAAGGAGA 1433
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TGCCATTCATAATTCATTCATTGCATCATGCACAGCCCAGAATGATACTACAAAAAGGAGA  960

Qy    1434 TCATCCAATTATCATGGGTTTTTGAACCCCCTTGTGAACCAGAGAGGCTACTTCCACTACCTT 1493
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TCATCCAATTATGATGGGTTTTTGAACCCCCTTGTGAACCAGAGAGGCTACTTCCACTACCTT 1020

Qy    1494 CCCTCGATATGCAAAAATAATTAAAATAATTGGTGAACTATATTTTGCAAGATTAAT 1553
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCCTCGATATGCAAAAATAATTAAAATAATTGGTGAACTATATTTTGCAAGATTAAT 1080

Qy    1554 AGAATAAGAATAAAAACTGTCTGTGAGTTGAATTAACAAATTTACATTGTAATCCTGGA 1613
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AGAATAAGAATAAAAACTGTCTGTGAGTTGAATTAACAAATTTACATTGTAATCCTGGA 1140

Qy    1614 TTTTTTTCTGTGAATTTAGTGAACAGTCACCATGTGTTCTTCTTCAGATTCTCGTGTTACAAAC 1673
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TTTTTTTCTGTGAATTTAGTGAACAGTCACCATGTGTTCTTCTTCAGATTCTCGTGTTACAAAC 1200

Qy    1674 CACTTTCCTGGTGGAGTAACAGTCATTGGAACTCATCATCTGTTGTTACTACAAGACATGGTGAA 1733
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 CACTTTCCTGGTGGAGTAACAGTCACCATGTGTTCATCTGTTGTTACTACAAGACATGGTGAA 1260

Qy    1734 AGATGCACTTCGGTCTCTTTGTCAGATCCTTCCCCAAGTGCTTACCTATAT 1793
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 AGATGCACTTCGGTCTTTTTGTCAGTCCTCCG--GTGCTTCCCCAAGTGCTTACCTATAT 1318

Qy    1794 AATAATCACCAGGACTAAGGACTGTATCGACACGAGACGATAAGCTTGGTCAT 1853
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1319 AATAATCACCAGGCATAACAGGGCTCGACACGAGACGATAAGCTTGGTCAT 1378

Qy    1854 TGTTAGTCTTATTCAGATCCATGGACGATAACAGGGCTGTGGTCAT 1913
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    1379 TGTTAGTCTTTATTCAGATCCATGGACGATAACAGGGTTGGTCAT 1438

Qy    1914 ATTCTTGAGGAATTTGCCACCTTGCAGGATG-------------AGGCAG 1973
           |||||||||||||||||||||||||||||||             ||||||
Db    1439 ATTCTTGAGGAATTTGCCACCTTGCAGGATG-------------AGGCAG 1475

Qy    1974 AGAAGGTTGAATGCAGCAGCGCCTTCACACATGCTGTTGAAACAGGAACCCAAAGGCAACATT 2033
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1476 AGAAGGTTGATGCAGCGCCTTCACACATGCTGTTGAAACAGGAACCCAAAGGCAACATT 1535

Qy    2034 TGGCCTGTTTAGGTACCTGGGGTCCTTTACCCATTAATACCTTACCATAAACCTTCGGCATTA 2093
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1536 TGGCCTGTTTAGGTACCTGGGTCCTTTACCCATTATGATACAGTACCTTC 1595

Qy    2094 TAAGTGGCTTTGAATTGGAACTCTACAGTATGCACGGAGTACTATTACATTATTGGTATC 2153
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1596 TAAGTGGCTTTGAATTGGAACTCTACAGTATGCACGGAGTACTATTACATTATTGGTATC 1655

Qy    2154 TCTCTGAATTCCTTTACGCATGGGTTGAATGTCAACATTGAGTCGTTCAACATTGAGTGTC 2213
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1656 TCTCTGAATTCCTTTACGCATGGGTTGAATGTCAACATTGAGTCGTCCGATGGCTCCAAA 1715

Qy    2214 TCGGCAGGAGAAAGGATAATGAAGAGCAGAGAAGGCCGTAGTAGTAAAAAACAAAGA 2273
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1716 TGGCAGGAGAAAGGATAATGGAAGAGCAGAGAAGGCCGTAGTAGTAAAAAACAAAGA 1775

Qy    2274 AAAAAAAAGAAAAGTTCGCCCATTGGACCCAGAGATTCACAATGAGCCAAGACATTACAGAACA 2333
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1776 AAAAAAAGAAAGGTTCGCCCATTGGACCCAGAGATTCACAATGAGCCAAGACATTCAGAACA 1835

Qy    2334 TGTGTGCTTGGAATGTTTAAAACCATGGTAGCATTTGACATTTGACATGGACCGCCAAAGTACGTAAAC 2393
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1836 TGTGTGCTTGGAATGTTTAAAACCATGGTAGCATTTGACATGGACGGCGAAAGTAGCGTAAAC 1895

Qy    2394 CGAAGTTTGAGCTTGATAGTGAACAAGTTCGGTATGAACACAGGTTTGCTCCATTCAACA 2453
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1896 CGAAGTTTGAGCTTGATAGTGAACAAGTTCGGTATGAACACAGGTTTGCTCCATTCAACA 1955

Qy    2454 GTGTGATGACCCCGCCGCCCAGTGCACTACTTACAGTTCAAGGAGAAATGTCTGACCTCAATA 2513
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1956 GTGTGATGACCCCGCCGCCCAGTGCACTACTTACAGTTCAAGGAGAAATGTCTGACCTCAATA 2015

Qy    2514 AATATAGCCCTCCTCCTCCCAGTCTCCTGAACTGTATGTGGCAGCACTTTCAAC 2573
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2016 AATATAGCCCTCCTCCTCCCAGTCTCCTGAACTGTATGTGGCAGCACTTTCAAC 2075

Qy    2574 AGGCAAAAATGATATTGGAAAATATTCCTAACCCGGACCAGAGGTTAATAGAATTTTAA 2633
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2076 AGGCAAAAATGATATTGGAAAATATTCCTAACCCGGACCAGAGGTTAATAGAATTTTAA 2135

Qy    2634 AGGTTGCCAAAACCCAACTTGTGGTTATGAAGTTATTGGCAGACACACAAAAAGGAAT 2693
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2136 AGGTTGCCAAAACCCAACTTGTGGTTATGAAGTTATTGGCAGACACACAAAAAGGAAT 2195

Qy    2694 CTAAAGTTCCTCCTGAATTTGATTTCTGTCATCATAATATATTTCCTGTTGTGAAAACTTG 2753
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2196 -CTAAAGTTCCTCCTGAATTTGATTTCTGCTCATTAAATATTTCCTGTTGTGAAAACTCT 2255

Qy    2754 TTTGAGAGAGACTGGGGAGGTGGGCCTATAAAGCGGGGCAGAGTCTTCTTCTTCAGACCCAACTCT 2813
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2256 TTTGAGAGAGACTGGGGAGGTGGGCCAGAAGGGCAGAGTCTTCTTCTTCAGACCCAACTCT 2315

Qy    2814 TAGAGGGGCACATCACCAGGCTCCACATCACGGGAAGTGAGATGGAATGGATTTCTTGGGTAACAA 2873
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2316 TAGAGGGGCACATCACCAGGCTCCACATCACGGGAAGTGAGATGGAATTTCTTGGGTAACAA 2375

Qy    2874 CTCATTAATAAGGAATACTTTTAGTTTGACAGGCCTTATATGACATGAATGAAAAACTGCTGT 2933
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2376 CTCATTATAAGGAAATACTTTTAGTTTGACAGGCCTTATATGACATGAATGAAAACTGCTGT 2435

Qy    2934 TTTAAAGTGGTTTATTATTGTTCCATGGAAGAAAAACTGGTCTTATTGAAT 2981
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2436 TTTAAAGTGGTTTATTATTGTTCCATGGTAAGACACTGGGTTCCATTAAT 2483
```

```
REFERENCE    1  (bases 1 to 2552)
  AUTHORS    Wenzlau,J.M. and Weiser-Evans,M.C.M.
  TITLE      Mouse Embryonic Growth Associated Protein (EGAP)
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 2552)
  AUTHORS    Wenzlau,J.M. and Weiser-Evans,M.C.M.
  TITLE      Direct Submission
  JOURNAL    Submitted (13-MAY-2002) Pediatrics, University of Colorado HSC,
             4200 East Ninth Ave, Denver, CO 80262, USA
FEATURES             Location/Qualifiers
     source          1..2552
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /dev_stage="embryo"
     CDS             55..2232
                     /codon_start=1
                     /product="embryonic growth-associated protein EGAP"
                     /protein_id="AAM52342.1"
                     /db_xref="GI:21539896"
                     /translation="MVWKATVDDDASGWELGVPEKMEKSSTSWWDITQDPEDACRELK
                     LGEILHDKLFGLFEAMSAIEWMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDLSL
                     PELIGIMDTCFCCLITWLEGHSLAQTVFTCLYIHNPDFIEDPAMKAFALGIKICDIA
                     REKVNKAAVEEEEDFQSMTYGFRMANSVTDLRVTGMLKDVEDDLQRRVKSTRSRQEE
                     RDPEVELEHQCLAAFSRVKFTRVLLTLLIAFTKKETSAVAEAQKLMVQAADLLSAIH
                     TSLHHGTQAQNGTTKGDHPIMMGFEPLVNQRLPFTPFRYAKIIKREEMVNYFSRLID
                     RIKTVCEVVNLPNIHCILDFFCEFSEQSPCVLSRSLLQTTFLVDKKVFGTHLMQDMV
                     KDALRSFVSPPVLSPKCCLYNNHQAKDCIDSFVTHCVRPFCSLVQTHGHNRARQRDKL
                     GHILEEFATLQDEAEKVDAALHTMLLKQEPQRHLACLGTWVLYHNLRIMIQYLLSGF
                     ELELYSMHEYYIYWYLSEFLYAWLMSTLSRADGSQMAEERIMEEQOKGRSSKKTKKK
                     KVRPLSREITMSQAYQNMCAGMERTMVAFDMDGKVRKPKFELDSEQVRYEHRFAPFN
                     SVMTPPVYHLQFKEMSDLNKYSPPPQSPELYVAASKHFQQTKMILENIPNPDREVSR
                     ILKVAKNPNVMKLLAGHKKESKVPPEFDFSVHKYFPVVKLV"
BASE COUNT      740 a    542 c    619 g    651 t
ORIGIN

Query Match    59.4%;  Score 1838.4;  DB 10;  Length 2552;
Best Local Similarity   84.4%;  Pred. No. 0;
Matches 2160;  Conservative   0;  Mismatches 341;  Indels 57;  Gaps  6;
```

```
Qy       1 GGAGAAGTAGGCATAAATGGTTATGGAAAGCTTCTTGTAGATGATGATGACGATGCAGAAGTGGGGAG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 GGAGAAGTAGGCATAAATGGTTATGGAAAGCTTCTTGTAGATGATGATGACGATGCAGAAGTGGGGAG 600

Qy      40 GGGCGCGGACGGCCATCATGGTTATGAAAAGCTACTGTAGAGACGATGCTTCGGGATGGGAG 99
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          GGGCGCGGACGGCCATCATGGTTATGAAAAGCTACTGTAGAGACGATGCTTCGGGATGGGAG 99
```

(sequence alignment continues — multiple Qy/Db rows of nucleotide sequence with alignment bars, numbered from 541/1 through 2160/2552)

RESULT 4
AF272892

LOCUS   AF272892   2543 bp   mRNA   linear   ROD 15-SEP-2000
DEFINITION Rattus norvegicus corneal wound healing related protein mRNA,
            complete cds.
ACCESSION   AF272892
VERSION   AF272892.1 GI:8926319
KEYWORDS   .
SOURCE   Rattus norvegicus (Norway rat)
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 2543)
  AUTHORS   Yi,X.J., Li,X.F. and Yu,F.S.
  TITLE   A novel epithelial wound-related gene is abundantly expressed in
            developing rat cornea and skin
  JOURNAL   Curr. Eye Res. 20 (5), 430-440 (2000)
   MEDLINE 20314365
   PUBMED   10855038
REFERENCE   2 (bases 1 to 2543)
  AUTHORS   Li,X. and Yu,F.
  TITLE   Direct Submission
  JOURNAL   Submitted (26-MAY-2000) The Schepens Eye Research, 20 Staniford
            Street, Boston, MA 02114, USA
FEATURES   Location/Qualifiers
     source   1..2543
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
     CDS   64..2241
                     /codon_start=1
                     /product="corneal wound healing related protein"
                     /protein_id="AAF81791.1"
                     /db_xref="GI:8926320"
                     /translation="MVMKAAVDDDASGWELNVPEKMEKSSTSWVDITQDPEDACRELK
                     LGELLHDKLFGLFERMSAIEMMDPKMDAGMIGNQVNRKVLNFEQAVKDGYIKIKDLSL
                     PELIGIMDTCFCCLITWLEGHSLAQTVFTCLYIHNPDFTEDPAMKAFALGILKICDIA
                     REKVNKAAAFEEEDFQSMTYGFKMANGVTDLRVTGMLKDVEDDMQRRVKSTRSRQGEE
                     RDPEVELEHQQCLAAFSRVKFRTVLITVLIAFTKKETSAVAEAQKLMVQAADLLSAIH
                     TSLHHGIQAQNGTTRKGDHPIMMGFEPLVNQRLLPPTFPRYAKIIKREEMVNYFSRLID
                     RIKTVCEVVNLPNLHCILDFFCEFSEQSPCVLSRSLLQTTFLVDNKKVFGTHLMQDMV
                     KDAPRSFVSPPVLSPKCCLYNNHQAKDICDSFVTHCVRPFCSLVQIHGHNRARQRDKL
                     GHILEEFATLQDEAREKVDAALHTMLLKQEPQRQHLACLGTWVLYHSLRIMQYLLSGF
                     DLELYSMHEYYIYWLSEFLYAWLMSTLSRADGSQMAAERIMEEQQKGRRSSKKTKKX
                     KKVRPLSRIITMSQAYQNMCAGMFKTMVAFDMDGKVRKPKFELDSEQVRYEHRFAPEN
                     SWTPPPVVHLQFKEMSDLSKYSPPQPPELVYAAS'GHQQAKMILESIPNADREVSR
                     ILKVAKPNFVVMKLLAGGHKKESKVPPEFDFSVHKYFPVKLV"
BASE COUNT   725 a   554 c   627 g   637 t
ORIGIN

Query Match   58.1%; Score 1799.2; DB 10; Length 2543;
Best Local Similarity   84.0%; Pred. No. 0;
Matches 2139; Conservative 0; Mismatches 343; Indels 64; Gaps 7;

Qy   550 GGCATAAATGGTTATGAAAAGCTTCTGTAGATGATGATGACGATTCAGGATGGGAGGCTCAGTATG 609
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   58 GGCATAAATGGTTATGAAAAGCCCCAGTAGATGATGACGACGCTTCGGGATGGGAGCCTCAATGTC 117

Qy   610 CCAGAAAAAAATGGAGAAAAAGCAATACAAACTGGGTGGGACAATTACCCAAGATTTTGAAGAA 669
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   118 CCTGAGAAAAAATGGAGAAAAAGTAGCACAAGCTGGGTGGGACAATTAACCCAGGACTTTGAAGAT 177

Qy   670 GCTTGTCGGAGAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTTTGAA 729
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178 GCTTGTCGAGACCTGAAGTTGGGAGAACTGCTTCACGATAAGCTGTTTGGTCTGTTTGAA 237

Qy   730 GCCATGTCTGCTATTGAAATGATGGAATCCCAAGATGGGAATCCTGGCATGATTGGAAAACCAA 789
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   238 GCCATGTCTGCTATTGAAATGATGGATCCCAAGATGGATGCGGGAATGATCGGGAACCAA 297

Qy   790 GTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTATTAAAATTAAA 849
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   298 GTGAATAGAAAAGTCCTCAATTTTGAACAAGCCCGTCAAGGATGGCACCATTAAAATTAAA 357

Qy   850 GATCTCACCTTGCCTCAAACTGATAGGGATTATGGATACATGTTTTTGCTGTTGATAACG 909
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   358 GACCTCAGCTTGCCTGAACTGATAGGAATTAAATGGATACCTGTTTTTGCTGTTGATCACA 417

...

Qy   1575 CTTTGAGCTGTGACAGCAATGCCATGATGAGCTACTACTACTACTGGTAGTACCTCCGA
          |||| ||||
Db   1634 ...

Qy   2161 ATTCCTTTTACGCAATGGTTGATGTCAACATTGAGTCGTGCCATGGCTCTCAAATGGCAGA 2220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1635 GTTCCTGTTGCAGGCTGATGTGTCAACCCTAGCCGCCGGCTGACGATGGCAGA 1694

Qy   2221 GGAAAGGATAAATGGAAGAGCAGCAGAAAGGCCGTAGTGTAGTATAAAAAAAAGAGAAATGA 2280
          ||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   1695 GGAGAGGATCATGGAGGAACAGCAGAAAGGCCGCAGCAAAAAAACAAGAAGAGAAAA 1754

Qy   2281 GAAAGTTCGCCCATTGAGCGCCGAGAGATCACAATGAGCCAAGCATATCACAATGTGTGC 2340
          ||||||||| ||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   1755 GAAAGTTCGTCCGTTGAGCCGGAGATCACAATGAGCCAGGCCTATCAGGAACATGTGTGC 1814

Qy   2341 TGGAAATGTTTAAAACCATGGTAGCATTTGACATGGACGGGCAAAGTACGTAAACCGAAGTT 2400
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1815 TGGGATGTTCAAGAACGATGGTTGGCTTTTGACATTGGCCAAAGTGGCCAAGCCAAGTT 1874

Qy   2401 TGAGCTTGATAGTGAACAAGTTCGGTATGAACACAAGTTGTTCCATTCAACAGTGTGAT 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1875 TGAGCTTGACAGTGAGCAGGTCCGGTATGAGCACAGGTTTGCTCCTGAGAAT 1934

Qy   2461 GACCCGCGCGCCAGTGCACTACTTCAGGTTCAAGGAGAAATGTCGGACCTCCAATAAAATATAG 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1935 GACGCCGCCACCAGTGCACTACCTGCGAGTTCAAGGAGAAATGTCTGACCTCCAATAAAATACAG 1994

Qy   2521 CCCTCCTCCTCAGTCCTGAAACTGTATGTGGCGAGCTAGTGAAGCTTTCAACAGGCAAA 2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1995 CCCTCCTCAGTCTCCAGTCGTCAGAGCTGTATGTGGCAGCTAGTGAAGCACTTTCAGCGAGCAAA 2054

Qy   2581 AATGATATTGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTAAAGGTTGC 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2055 AATGATCCTTGAAAAATATCCCCAATCCAGAGGTCAGTAGAATCCGGAAGGTTGC 2114

Qy   2641 CAAACCCAACTTTGTTGGTTATGAAGTTATTGGCAGGAGCACAACAAAAAGGAATCTAAAGT 2700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2115 CAAGCCCAACTTTGTTGGTTATGAAGCTCTTAGCGAGGAGCACAAAAAAGAGTCAAAAGGT 2174

Qy   2701 TCCTCCTGAATTTGATTTCTCTGCTCATAAATATTTTCCTGTTGTGTGAAACTTGTTGTTGAGA 2760
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2175 TCCTCCCGAATTCGACTTCTCCGTTCACAAATATTTTCCTGTTGTGAAACTGGTTTGAAA 2234

Qy   2761 GAGACTGGGGAGGTGGCCATAAAAGGGGCCAGAGTCTTCTTTCAGACCCAACTCTTAGAGGG 2820
          |||
Db   2235 GAGACTGCAAAGATGACCGTGTCCACTC------------TAGGGA 2269

Qy   2821 CACATCACCAGGCTCCACATCACG--GGAAGTGAGATGGATTCTTGGGTAACAACTCAT 2878
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   2270 GACATCACGTCTGCCACACAGGTGACAAAGTGGATCTCTGAATGACAGACAGCTTGC 2329

Qy   2879 TATAAGGAATACTTTTAGTTTGACAGCCTTATATGACAATGAAAAAACTGCTGTTTTTAA 2938
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2330 TACAGGAAGACTTTTCAGTTTACCACCTTACCTAGAGAGAGAATGAGAACTGCTGTTTTAA 2389

Qy   2939 AGTGGTTTATTATTATGTTCCATGGAAGAAAACTGTCTTTATTGACATGATGAACGTTTAT 2998
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2390 AGIGGTTTATTGTACTCCCATGGGTGTAACTGG---CTGGAATCCATTGGGTAAATGGATCTAC 2445

Qy   2999 ATGGTTTTATTTAATTACAGAATTTAATCACAAAATCATTTTATGTTGAAATAGTAGT 3058
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2446 ATGGTTTGATTACAGACTTAATTGTAAACAATTTTTAATGAAAGATTGTGTG-AAATAGT 2504

Qy   3059 GTTTATAAAGGTTAATAAATTTCTTGACAAAAAAAAA 3096
          |||||||||||||||||||||||||||||||||||||
Db   2505 GTTTGTAAAGGTTAATAAATTTCTTGACAAAAAAAAA 2542

RESULT 5

Db   AK025266

```
Db   2518  AAGGTTAATAAATTTCTTGACAAAA 2543
           ||||||||||||||||||||||||
```

LOCUS   AK025266   1839 bp   mRNA   linear   PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ21613 fis, clone COL07381.
ACCESSION   AK025266
VERSION   AK025266.1 GI:10437744
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1839)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES   Location/Qualifiers
     source   1..1839
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="COL07381"
                     /tissue_type="colon"
                     /clone_lib="COL"
                     /note="cloning vector pME18SFL3"
     CDS   134..1714
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15097.1"
                     /db_xref="GI:10437745"
                     /translation="MVWKASVDDDSGWELSMPEKMEKSNTNWVDITQDFEEACRELK
                     LGELLHDKLFGLFEAMSAIEMMDPKMDAGMIGNQVNRKVLNFEQAIKDGTIKIKDITL
                     PELIGIMDTCFCCLITWLEGHSLAQTVFPCLYIHNPDFIEDPAMKAFALGILKICDIA
                     REKVNKAAVFEEEDFQSMTYGFRMANSVTDLRVTGMLKDVEDDMQRRVKSTRSROGEE
                     RDPEVELEHQRCLAVFSRVKFTRVLLTVIAIFTKKETSAVAEAQKLMVQAADLLSAIH
                     NSLHHG1QAQNDTTKGDHPIMVGFEPLVNQRLLPPTFPRYAKIIKREEMVNYFARLID
                     RIKTVGEVNLTNLHCILDFFCEFESEQSPCVLSRSLLQTTFELVDNKKVFGTHLMQDMV
                     KDALRSFVSPPVLSPKCYLYNHQAKDIIDSFVTHCVRFPCSLIQIHGHNRARQRDKL
                     GHILEEFATLQDEAEKVDAALHTMLLKQPEPQRQHLAWLGTWVLYHNLRIMIQYLLSGF
                     ELELYSMHEILLHILVSL"
```

BASE COUNT      565 a    336 c    437 g    501 t
ORIGIN

```
Qy    Query Match   57.1%;  Score 1766.6;  DB 9;  Length 1839;
      Best Local Similarity  98.1%;  Pred. No. 0;
      Matches 1828; Conservative 0; Mismatches 9; Indels 26; Gaps 3;

Qy    423 ATACGCANGCGTCCACGCGTCGCCCGGTCGGCGTCGAGGAGGGAAGGGGGGCGCGGGCGGC 482
          ||||||||||||||||||||||||||||||||||| ||||||||||||||| ||||||||
Db      1 ATACGCANGCGTGCCACGCGTTGCCCGGGTTCGGGGTTGAGAGGGAGGGGGGAGGGGGGC 60

Qy    483 CGAGGCGGCGGTCGTTATTTCCGTGGTCCGACAGTGGGCGGGGGCGGCGGGGGCCGCACGGG
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGAGGCGGCGGGTCGTTATTTCCGTGGTCCGACAGTGGGCGGGGGCGGCGGGGGTGACCACGG 120

Qy    543 AGAAGTAGGCCATAATAATGGTTATGAAAGCTTCTTGTTAGATGATGACGATTGGGAGCT 602
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGAAGTAGGCCATAATAATGGTTATGAAAGCTTCTTGTTAGATGATGACGATTGGGAGCT 180

Qy    603 CAGTAGTGCCAGAAAAATGGAGAAAAGCAATACAAACTGGGGTGGACATTACCCAAGATTT 662
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAGTATGCCAGAAAAATGGAGAAAAGCAATACAAACTGGGGTGGACATTACCCAAGATTT 240

Qy    663 TGAAGAAGCTTGTCAGAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTGGTCT 722
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGAAGAAGCTTGTGCAGAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTGGTCT 300

Qy    723 TTTTGAAGCCATGTCTGCTATTGAAATGATGGATCCCAAGATGATGTCTGAAGATTGG 782
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TTTTGAAGCCATGTCTGTTATTGAAATGATGGATCCCAAGATGATGTCTGAAGATTGG 360

Qy    783 AAACCAAGTTAAATGAAAAAGTTCTCAAATTTGAACAAGCTATCAAGGATGGCACTATTAA 842
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAACCAAGTTAATGAAAAAGTTCTCAAATTTGAACAAGCTATCAAGGATGGCACTATTAA 420

Qy    843 AATTAAAAGATCTCACCTTGCCTGAACTGATAGGGGATTATGGATACATGTTTTTGCTGTT 902
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AATTAAAAGATCTCACCTTGCCTGAACTGATAGGGGATTATGGATACATGTTTTTGCTGTT 480

Qy    903 GATAACGTGGTTAGAAGGCCATTCACTGCACAGACAGTATTTACGTGCCTTTACATTCA 962
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GATAACGTGGTTAGAAGGCCATTCACTGCACAGACAGTATTTACGTGCCTTTACATTCA 540

Qy   1023 CTGTGACATTGCAAGGGAAAAAAGTAAATAAAAGCTGTCGTTGTTTTTGAAGAGGAGAATT 1082
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TAATCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAAAT 600

Qy    963 TAATCCAGACTTTATAGAAGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCTTGAAAAT 1022
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTGTGACATTGCAAGGGAAAAAAGTAAATAAAAGCTGTCGTTGTTTTTGAAGAGGAGAATT 660

Qy   1083 GTCAATGACTTATGGATTTAAAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCAT 1142
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GTCAATGACTTATGGATTTAAAAATGGCTAACAGTGTGACAGATCTTCGAGTTACAGGCAT 720

Qy   1143 GCTAAAAGATGTGGAGGATCACATGCAAAGAAGAGAGTAAAGCAGTACTCGACAAGG 1202
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GCTAAAAGATGTGGAGGATCACATGCAAAGAAGAGAGTAAAGCAGTACTCGACAAGG 780

Qy   1203 AGAAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAAATGTTTAGCAGTATTCAGCAG 1262
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AGAAGAAAGAGATCCAGAAGTTGAACTAGAACACCAACAAATGTTTAGCAGTATTCAGCAG 840

Qy   1263 AGTGAAATTACTCGGTCTGTGTTACTAGGCCTTTACTAAGAAGAAAAGAGAGACCAG 1322
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AGTGAAATTACTCGGTCTGTGTTACTAGGCCTTTACTAAGAAGAAAAGAGAGACCAG 900

Qy   1323 TGCTGTTGCAGAAGCTCAAAAAATTGATGGTTCAAGCAGCAGAGACATTCTTTCTTGCATTCA 1382
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TGCTGTTGCAGAAGCTCAAAAAATTGATGGTTCAAGCAGCAGAGATCTTCTTTCTGCCATTCA 960

Qy   1383 TAATTCATTGCATCATGGCATCCAGCCCCCAGAATGATACTACAAAAGGAGAGCATCCAAT 1442
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TAATTCATTGCATCATGGCATCCAGCCCCCAGAATGATACTACAAAAGGAGAGCATCCAAT 1020

Qy   1443 TATGGAGGTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACTACCTCCCTCGATA 1502
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 TATATGGGGTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACTACCTCCCTCGATA 1080

Qy   1503 TGCAAAAAATAATTAAAAAGGAAGAAAATGGTGAACTATTTTTGCAAGATTAATAGATAGAAT 1562
          ||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   1081 TGCAAAAATAATTAAAAAGGGAAGAAAATGGTGAACTATTTTTGCAAGATTAATAGATAGAAT 1140

Qy   1563 AAAAAACTGTCTGTGAGGTTGTGAATTTAACAAAATTTACATTGTATCCTGGATTTTTTCTG 1622
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
RESULT 6
BC053286
BC053286

Qy  1141 AAAAACTGTCTGAGTTGTGAGGTTGTGAATTAACAAATTAACATTGTATACCTGGATTTTTCTG 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1623 TGAATTTAGTGAACAGTGACCACCATGTGTTCTTCTTCAAGATCTCGTTTACAAACCACTTTCCT 1682

Qy  1201 TGAATTTAGTGAACAGTGACCACCATGTGTTCTTCTTCAAGATCTCGTTTACAAACCACTTTCCT 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1683 GGTGGATAACAAAAAGTCTTTGGAACTCATCTCATGCAAGACACATGGTGAAAGATGCACT 1742

Qy  1261 GGTGGATAACAAAAAGTCTTTGGAACTCATCATCATGCAAGACACATGGTGAAAGATGCACT 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1743 TCGGTCTTTTGTCAGATCCTCCCAAGTGCTACCCTATATAATAATCAC 1802

Qy  1321 TCGGTCTTTTGTCAG--TCCTCCGGTGCTTCCCCCAAGTGCCTACCCTATATAATAATCAC 1378
         ||||||||||||||      ||||||||||||||||||||||||||||||||||||||||
Db  1803 CAGGCTAAGGACTGTATCGACTCCTTTGTTGTTACTCACTGTGTTTCGGCGCCATTCTTCAGTTCTT 1862

Qy  1379 CAGGCTAAGGACTGTATCGACTCCTTTGTTGTTACTCACTGTGTTTCGGCGCCATTCTTCAGTTCTT 1438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1863 ATTCAGATCACCATGGGACATAACACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAG 1922

Qy  1439 ATTCAGATCACCATGGGACATAACAGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTGAG 1498
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1923 GAATTTGCCACCTTGCCAGGATGAGTTATGACATTTATTTAATAGGCAGAGAAGGTTG 1982

Qy  1499 GAATTTGCCACCTTGCCAGGATG-----------AGGCAGAGAAGGTTG 1535
         ||||||||||||||||||||||          ||||||||||||||
Db  1983 ATGCACGCGCTTCACACACATGCTCTTTGAAACAGGAACCCCAAAGGCAACATTTGGCCTGTT 2042

Qy  1536 ATGCACGCGCTTCACACACATGCTCTTTGAAACAGGAACCCCAAAGGCAACATTTGGCCTGTT 1595
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2043 TAGGTACCTGGGGTCCTTTACCATAACCTTCGCAGTTGTACCTTCTCAAAGTGGCT 2102

Qy  1596 TAGGTACCTGGGGTCCTTTTACCATAACCTTCGCATTATACAGTTGTACCTTCAAGTGGCT 1655
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2103 TTGAATTGGAACTCTCACGTATGCACGAG-TACTATTACATTATATATTGGTATGTCTCTGAA 2161

Qy  1656 TTGAATTGGAACTCTCACAGTATGCACGAGATACTATTACATTATATATTGGTATCTCTGAA 1715
         ||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db  2162 TTCCTTTACGCATGGTTGATTGATTGAATGTCGTGCCGATGGTCTCAAATGGCAGAG 2221

Qy  1716 TTCCTTTACGGCATGGTTGATTGGTCAACATGATTGAGTCGGTGCCGATGGTCTCAAATGGCAGAG 1775
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2222 GAAAGGATAATGGAAGAGACAGCAGCAGAAAAGGCCGTAGTAGTTAAAAAACAAAGAGAGAG 2281

Qy  1776 GAAAGGATAATGGAAGAGGAGCGAGCAGAGAAGGCCGTAGTAGTTAAAAAACAAAGAGAGAG 1835
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2282 AAA 2284

Qy  2282 AAA 2284
         |||
Db  1836 AAA 1838
```

LOCUS   BC053286   2467 bp   mRNA   linear   VRT 09-JUN-2003
DEFINITION Danio rerio cDNA clone MGC:64157 IMAGE:6797204, complete cds.
ACCESSION   BC053286
VERSION   BC053286.1 GI:31418975
KEYWORDS   MGC.
SOURCE   Danio rerio (zebrafish)
  ORGANISM Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 2467)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2 (bases 1 to 2467)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-JUN-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Leonard I. Zon, M.D.
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site: http://www-shgc.stanford.edu
            Contact: (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, GenomeScan gene prediction, Similarity but not identity
            to protein.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 117 Row: n Column: 11
FEATURES   Location/Qualifiers
     source   1..2467
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="MGC:64157 IMAGE:6797204"
                     /tissue_type="Kidney, zebrafish"
                     /clone_lib="NCI_CGAP_Zkidi1"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6.1"
     CDS   58..2232
                     /codon_start=1
                     /product="Unknown (protein for MGC:64157)"
                     /protein_id="AAH53286.1"
                     /db_xref="GI:31418976"
                     /translation="MVMKSSVEEEEGGWGLGIPEKMRNNANWVDVTQEFKGACKELKL
                     GELLHDKLFGLFEAMSAIEMDPNDAGMIGNQVNRKVLNFDQAVKDEAIRVRDLSIP
                     ELIGIMDTCFCCLITWIEGHSLAQTVFTCLYVHPNDLIQDPALKAFALGILKICDIAR
                     EKVNKAAVFEEEDFQAMTYGFRVWFTRLLLSALISFTKKETSAVSEAQKLMSQAADLPAVHA
                     DPEVELDHQQCLALFSRVKFTRLLLSALISFTKKETSAVSEAQKLMSQAADLPAVHA
                     TIOYGIOSONDTKGDHPIMMGFEPLVNQRLLPPTFPRYARKIIKREEMVNYFSKLIER
                     IKSVCEVININLHSIDIFHCEFSEQSPCVLSRSILCTTFLIDNKVFGTHLMQDMIK
                     DALRCFVSPPVLSSKCSLNNNHQAKDYIDSFVTHCTRPFGSLIGHGNRAQRDKLG
                     HILEEFATLQDEAEKVDAALHGLLMKLEPQRQHLACLGTWILLYHNLRIMIQVLLSGFE
                     LELYSMHEYYIYWYLSEFLYAWMLMSTLSRADSSQMAEBERILEEQLKVRSSKKSKKK
                     KARPLSXEITMSQAYQNMCAGMXTMIALDMDRKVRKPQFELDSEQVRYERPAPFNS
                     VVTPPPVHYIQPKEMSDLKKYNPPRSADLYMAASKHFQQAKLLLENVTSPDAEVNRI
                     LKVAXPNIVVMKLLAGGHKKEFKALPEFDFSAHKYFPIVKLL"
BASE COUNT   720 a   574 c   600 g   573 t
ORIGIN

Query Match   40.9%;   Score 1264.8;   DB 5;   Length 2467;
```

Best Local Similarity   74.3%;   Pred. No. 8.3e-276;
Matches 1648;   Conservative   0;   Mismatches 542;   Indels   28;   Gaps   3;

```
Qy    543  AGAAGTAGGCAATAATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATTGGGAGCT    602
           |||||||||||||||  ||||  ||| ||||||  || ||||| || | | ||| ||||||||
Db     45  AAAAGCTGTCTCCATGGGTGATGAAGTCATCGGTCCGAGGAGGAGGAGGATGGGGTCT    104

Qy    603  CAGTATGCCAGAGAAAAAATGGAGAAAAGCAATACAAAACTGGGTGGACATTACCCAAGATTT    662
Db    105  TGGGATCCCAGAGAGAT---GAGAACAATGCCAACTGGGTGGACCTCACCCAGGAGTT    161

Qy    663  TGAAGAAGCTTGTGTGCAGGAATTAAAGTTGGGAGAATACTTACCTTCATGATAAGAGCTT    722
Db    162  TAAAGGAGCTTGTAAAGAACTCAAACTCGGGCGAGTCGCCCATGACAAGCGTATTCGGTCT    221

Qy    723  TTTTGAAGCCATGTTGTGATTGAATGATCAACATTGCATGATGGTATGGATGCTGGCATGATTGGTCT    782
Db    222  GTTCGAAGGCCATGTCCGCCATTGAGATGATGGATGCTGGAATGATGGTTCT    281

Qy    783  AAACCAAGTTAATGAAAAGTTCTCAAATTTGAACAAGCTATCAAGGATGGCACTATTAA    842
Db    282  AAAATCAAGTCAACCGTAAAGTCCTCAACTTTGACCAAGCTGTTAAGGATGAAGCCATACG    341

Qy    843  AATTAAAGATCTCACCTTGCCTGAACTGATAGGATAGGAGGACACATGTTTCTGCTGTTT    902
Db    342  GGTGAAAGATCTAAGTATTCCTGAGCTCATCGGGATCGGGATTATGGATACACATGTTTCTGCTGTTT    401

Qy    903  GATAACGTTGGTTAGAAGGCCATTCACTGGCACAGACAGTAGTGCCTTTACATTCA    962
Db    402  GATCACGTGGCTGGGAGGGGGCATTCTTGGGTTCAGACCGTTCACATGTTCTACGTTCA    461

Qy    963  TAATCCAGACTTTATAGAAGAATCCTGCTATGTGAAGGCTTTTGGTCTTGGGAAATCTTGAAAAAT    1022
Db    462  TAATCCCGGATCTGATCCTGGCCTCAAAGCTTCGGCTTCTGGGGCATCCTGAAAAAT    521

Qy   1023  CTTGTGACATTGCAAGGGAAAAAAGTAAATAAAGCTGCTGTTGTTTTTGAAGAGGAAGAT    1082
Db    522  CTGTGACATCGCCGCCTGTAAAGGTCATGGGATCTTGTTTTGAAGGAGGAAGAGATTTTCA    581

Qy   1083  GTCAATGACTTATGGATTTAAAATGGCTAACAGTGTGAACAGAGTTCGGAGTTTACAGGCAT    1142
Db    582  GGCCATGACTTACGGGCTTTTAAAATGGCTGACAACAAGCTGCCACAACACGGGGAAT    641

Qy   1143  GCTAAAAGAGTGTGGGAGGAGATGACATGCAAAGAAGAGAGTAGAGTACTCGACAAGTCGACAAGG    1202
Db    642  GCTGAAAGAGATGTTGAGGGAATGAACTTCAGAGGAAAGTAAAGAGTACCCGCAGTCGCACAGGG    701

Qy   1203  AGAAGAAGAGATCCAGAGAGTTGAACTAGGACACCACAAATGTTTAGCAGTATTCAGCAG    1262
Db    702  TGAACAGCCAGATCCTGAGGTGGAACTGGATCATCAGCAGTAGTGGCTCTTGGTTCAGTCG    761

Qy   1263  AGTGAAAATTTACTCGTGTGTTACTGACAGTGGCTTATAGCCTTTACTTAAGAAAGACCAG    1322
Db    762  GGTTAAATTCACTCGTCTGTGCTGCGCCTAAATCCTTCACCAAAAAAGAGACGAG    821

Qy   1323  TGCTGTTGCAGAAGGTCAAAAATTGATGGTTCAAGCAGACAGATCTTCTTTCTGCCATTCA    1382
Db    822  TGCTGTGAGCGGAAGCACAGAGCTGATTGAGTGAAGCTGGCCTGAGTCCA    881

Qy   1383  TAATTCAATTGCATGGATCCAGCCCAGAAATGATACTACAAAAAGAGATCATCCAAT    1442
Db    882  TGCTACCATACAGTACGGCCACATCAGGTGCAGAACGACACTACCAAGGAGATCACCCGAT    941

Qy   1443  TATGATGGGGTTTTGAACCCCCTTGTGAACCAGAGGCTACTTCCACCTACCTTCCCTTCGATA    1502
Db    942  CATGATGGGGTTTTTGAGCCCTCGGTTAACCACGCGCCCTGCTCCACTACTGTCCCCGATA    1001

Qy   1503  TGCAAAAATATTAATTAAAGGGGAAGAAAGGGTGAACTATTTGCAAGATTAATAGATGAAT    1562
Db   1002  TGCCAAGATCATCAAGCGACAGAAGAGGTGGTGAAGCTGACGAGAGGATGGCTCAAGCTC    1061

Qy   1563  AAAAACTGTCTGTGAGGTTGTGAATTTAACAAATTTACATTGTATCCCTGGATTTTTTCTG    1622
Db   1062  TGCCAAGAATTTCAGCGAAAACTCATTCACTTGAGCGCAT    1061
```

```
Db   1062  TAAGAGCCGTGGTGTGTGGAGGTCATCAACACATCAACCTGACCAACCTGCCACAGCATCCCTGGATTTCTTCTG    1121
Qy   1623  TGAATTTAGTGAACAGTCACCAGTGTGTTCTTTCCAAGATTCTCTTACAAACCACTTCCT    1682
Db   1122  TGAATTCAGTGAGCAGTCCCATGTGTCCTTTCTAGAATCTTCACTGCACACGTTTCT    1181

Qy   1683  GGTGGATAACAACAAAAGGTCTTTGGAACTCATGCATGCTTCAGAAGCACATGGTGAAAGATGCACT    1742
Db   1182  GATGGACAACAAGAGAGGTGTTTGGGAACTCAATGCCAGGATATGAAAGATGGCT    1241

Qy   1743  TCGGTCTCTTTTGTCAGATCCTCCGAGTGCTTCCCCGAGTGCTACCTATATAAATCAC    1802
Db   1242  GCGCTGTTTGTCAG--TCCACCGTTCTTTCCTCAAAATGCAGTTTGAATAAATAAACCAC    1299

Qy   1803  CAGGCTAAGGACTGTATCGACAATAACAGGGGGCTCGACAGAGAGATAAGCTTGGTCAATTCTTGTCAGAG    1862
Db   1300  ATCCAGATTCACGGACACAACCGAGCCCGACAGAGGTCAAATCTGGGGTCACATTCTGGAG    1359

Qy   1863  ATTCAGATCCATGGACATAACAGGGGCTCGACAGAGAGATAAGCTTGGTCATATTCTTCTTGAG    1922
Db   1360  TGGGCACCCTGGGATCCTCTTACCACCAACAACCTGCGCATCATGACTCAGTCAGGGTTGGAG    1419

Qy   1923  GAAATTTGCCACCTTGCCATGGAGGATGAGTTTATGACATTTTTAATTAATTTTTTGCTGTTTT    1982
Db   1420  GAGTTCGCCACACTTCAGGATG------------------AGGCAGAGGTGG    1456

Qy   1983  ATGCAGCGCGTTTCACACCATGCTGTTGAAACAGGAACCCCAAAGGCAACATTTGGCCGTT    2042
Db   1457  ACGCTGCTCTTCACGGTCTGCTGATGAAGCTGGAGTCGGGAGCGCAGCGCATTTAGCCGTC    1516

Qy   2043  TAGGTACCTGGGTCTTCCTTTACCATAACCTTCGCATTATGATACAGTACTTCTAAGTGGCT    2102
Db   1517  TGGGCACCTGGATCCTCTACCACAACCTGCGCATCAGTGATGCCAGTCAGGCTCAGGCGGCT    1576

Qy   2103  TTGAATTGGAACCTACAGTATGCACAGAGTACTGTTACTGCTATTACATATATTGGTATTCTCTGAAT    2162
Db   1577  TTGAGCTGGAGCTCACAGCATGCACAGAATACTATTACATCGGTCCTGTCTCAGAGT    1636

Qy   2163  TCCTTTACGCATGGGTTGATGTGCAACATTGTTGAAACAGGACCCCAAAGGCAACATTTGGCCGTT    2222
Db   1637  TCTTGTATGCGCTGGCTGATGTGGACATGCACGAATACTATTACAATCCATGGCAAGGCTTC    1696

Qy   2223  AAAGGAATAATGGAAGAGCAGCAGAGAAAGGCCGTAGTAGTGAAAAACAAAGGAAAAAAAAGA    2282
Db   1697  AGAGAAATCCTGGAGGAAGCAGTCAAAAGTGCCAAGCAAGAGAGAGAGAGA    1756

Qy   2283  AAGTTCGCCCATTGAGGCGAGAGATCACAATGAGCCAAGCACATATCAGAAGCAATGTGTGCTG    2342
Db   1757  AAGCCTCGTCCTCTGAGTAAAGAGATCACACCATGAGCCCATGACCATTGTGTCG    1816

Qy   2343  GAAATGTTTAAAACCATAGGTAGCATTTTGACATGATTGGCGGGCAGGTAAACGGAAGTTTTG    2402
Db   1817  GGATGTTACAAGGACGATGATCGGGCTGGATATATGGGCGCAAAACCTCAGTTTG    1876

Qy   2403  AGCTTGGATAGTAGTGGAACAAGTTCGGTATGAACACAGGTTTGCTCCATTCAAACAGTGTGATGA    2462
Db   1877  AGCTGGACAGTGAGCAGGTGCCCTACCAGCATGCCGTTTCTCAACAGTGGTGTCA    1936

Qy   2463  CCCCGCGCCCAGTGCACTACTTACAGTTCAGAGGAAAGTGTCGACCTCAATAAAATATAGCC    2522
Db   1937  CACCACCACCAGTTCACTACTACATCCAGTTTAAGGAAATGTCCGGATTTGAAGAAGTCAATC    1996

Qy   2523  CTCCTCCTCAGTCTCCTGAACTGTTATTTTGGGCACGCTTCACACGGCAAAAA    2582
Db   1997  CTCCTCCTCGCTCGACCGGATCTCCTATATGGCCGCAGTGAAACATTTCCAGCAGGCCAAAAC    2056

Qy   2583  TGATATTGGAAAAATATTCCTAACCCCGGACCATGAGGTTAATAGAATTTTAAAGGTTGCCA    2642
Db   2057  TCTTACTGGAAAATGTCACCAGCCCCTGATGCGAGGTTAATCGGTTATCCTTAAAGTGGCCA    2116

Qy   2643  AACCCAAACTTTGTGGGTTATGAAGTTATTGGCAGGAGGCACAAAAGAATCTAAAGTTC    2702
Db   2117  AACCCAATATTGTTGTTAATGAAGCTGCTGGCTGGAGGAGGACAACAAGAGACCAGGCCC    2176
```

RESULT 7

```
Qy   2703  CTCCTGAATTTGATTTCTCTGGCTCATAAATATTTCCTGTTGTGAAACTTGTTGAGA  2760
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2177  TTCCTGAATTTGACTCTCTGCTTCACAAATACTTCCCATAGTCAAAATCCTATGAGA  2234
```

LOCUS   AK026296   997 bp   mRNA   linear   PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ22643 fis, clone HSI07031.
ACCESSION   AK026296
VERSION   AK026296.1 GI:10439112
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
  AUTHORS   Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
            Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
            Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 997)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan;cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction; 5'- & 3'-end one pass sequencing: Department of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES   Location/Qualifiers
     source   1..997
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HSI07031"
                     /tissue_type="human small intestine"
                     /clone_lib="HSI"
                     /note="cloning vector pME18SFL3"
     CDS   15..650
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB15435.1"
                     /db_xref="GI:10439113"
                     /translation="MHEYYIYWYLSEFLYAWLMSTLSRADGSQMAERRIMEEQQKGR
                     SSKTKKKKVRPLSREITMSQAYQNNCAGMFKTMVAFDMDGKVRKPKFELDSEQVRY
                     EHRFAPFNSVMTPPPVPHYLQFEKEMSDLNKYSPPPQSPELYVAASKHFQQAKMILENIP
                     NPDHEVNRILKVAKENFVVMKLLAGGHKKESKVPEFDFSAHXFPVXLIv"
BASE COUNT   334 a   171 c   213 g   279 t
ORIGIN
```

Query Match             31.9%;  Score 988;  DB 9;  Length 997;
Best Local Similarity   100.0%;  Pred. No. 4.4e-213;
Matches  988;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

```
Qy   2109  TGGAACTCTACAGTATGCACAGTACTATTACATATATATTGGTATCTCTCGAATTCCTTT  2168
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  TGGAACTCTACAGTATGCACAGTACTATTACATATATATTGGTATCTCTCGAATTCCTTT    60

Qy   2169  ACGCATGGTTGATGTCAACATTGAGTGTCGTGCCGATGGCTCTCAAATGGCAGAGGAAAGGA  2228
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  ACGCATGGTTGATGTCAACATTGAGTGTCGTGCCGATGGCTCTCAAATGGCAGAGGA    120

Qy   2229  TAATGGAGGAGCAGCAGAAAAGGCCGTAGTAGTAAAAAAACAAAGGAAGGAAAGTTC  2288
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  TAATGGAGGAGCAGCAGAAAAGGCCCTAGTAGTAAAAAAAACAAAGGAAAGAAAGTTC   180

Qy   2289  GCCCATGAGCCGGAGAGATCACATGAGCCAAGCAATATCACAACACATGTGTGGAAATGT  2348
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    181  GCCCAGTGGACCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAAATATAGCCTC  240

Qy   2349  TTAAAACCATGGTAGCATTTGACATTGGACGGCCAAAGTACGTAAACCGAAGTTGAGCTTG  2408
           ||||||||||||||||||||||||||||||||||||||||||||
Db    241  TTAAAACCATGGTAGCATTTGACATTGGACGGCCAAAGTACGTAAACCGAAGTTG    300

Qy   2409  ATAGTGAACAAGTTCGGTATGAACACAGGTTGCTCCATTCAACAGTGTGATGACCCCGC  2468
           ||| |||||||| ||||||||||||||||||||||||||| |||||||
Db    301  ATAGTGAACAAGTTCGGTATGAACACAGGTTTGCTCCATTCAACAGTGTGATGACCCCGC   360

Qy   2469  CGCCAGTGCACTACATTCACAGTTCAAGGAAATGTCTGACCTCAATAAAATATAGCCCTCTC  2528
           ||||||||| ||| ||||| |||||||||||||||||||||||| ||||
Db    361  CGCCAGTGGACCACTACTTACAGTTCAAGGAAATGTCTGACCTCAATAAAATATAGCCCTCTC  420

Qy   2529  CTCAGTTCTCCTGAACTCTGTATGTCGGCAGCTAGTAAGCAACTTCAAACAGCGAAAATGATAT  2588
           ||||||||||||||||||||||||||||||||||||||||||||
Db    421  CTCAGTCTCCTGAACTCTGTATGTGGCAGCTAGTAAGCAACTTCAAACAGCGAAAATGATAT   480

Qy   2589  TGGAAAATAATTCCTAACCCGGACCATGAGGTTAATAGAATTTAAAGGTTGCCAAACCCA  2648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  TGGAAAATAATTCCTAACCCGGACCATGAGGTTAATAGAATTTAAAGGTTGCCAAACCCA   540

Qy   2649  ACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAAGGAAATCTAAAGTTCCTG  2708
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  ACTTTGTGGTTATGAAGTTATTGGCAGGAGGACACAAAAAAGGAAATCTAAAGTTCCTCCTG  600

Qy   2709  AATTTGATTTCTCTGCTCATAAATATTTCCTGTTGTGTGAAACTTGTTGAGAGAGACTGG  2768
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  AATTTGATTTCTCTGCTCATAAATATTTCCTGTTGTGTTGAAACTTGTTGAGAGAGACTGG   660

Qy   2769  GGAGGTGGCCAATAAAGGGGGCAGAGTCTTCTTCAGACCCAACTCTTAGAGGGGACATCAC  2828
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  GGAGGTGGCCAATAAAGGGGGCCAGAGTCTTCTTCAGACCCAACTCTTAGAGGGGACATCAC  720

Qy   2829  CAGGGCTCCACATCACGGGAGAGTGAGATGAGATTTCTTGGGTAACAACTCATTATAAGGAAT  2888
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  CAGGGCTCCACATCACGGGAAGTGAGATGGGATTTCTTGGGTAACA-CTCATTATATAAGGAAT  780

Qy   2889  ACTTTTAGTTTGACAGCCTTATATATGACAGTAAATGAAAAACTGCTGTTTTGTTTAAAGTGGTTTAT  2948
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  ACTTTTAGTTTGACAGCCTTATATGACAGTAAATGAAAAACTGCTGTTTTGTTTAAAGTGGTTTAT   840

Qy   2949  TATGTTCCATGGAAGAAAACTGGTCTTATTGAAATGCATTGATGAAACGTTATATGGTTTTAT  3008
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  TATGTTCCATGGAAGAAAACTGGTCTTATTGAAATGCATTGATGAAACGTTATATGGTTTTAT  900

Qy   3009  TACAGATTTAATCACAAATCATTTTTTATGAATGATGAGTGAAAAATAGTGTTTATTAAAG  3068
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  TACAGATTTAATCACAAATCATTTTTTATGAATGATGAGTGAAAAATAGTGTTTATTAAAG  960

Qy   3069  GTTAATAAATTTCTTGACAAAAAAAAAAA  3096
           ||||||||||||||||||||||
Db    961  GTTAATAAATTTCTTGACAAAAAAAAAA  988
```

RESULT 8

```
LOCUS       AK056059               1796 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ31497 fis, clone NT2NE2005317, highly similar
            to Rattus norvegicus corneal wound healing related protein mRNA.
ACCESSION   AK056059
VERSION     AK056059.1  GI:16551204
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
```

AUTHORS   Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
           Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
           Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
           Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
           Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
           Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
           Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
           Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
           Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
TITLE   NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 1796)
AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
TITLE   Direct Submission
JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
           Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
           (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3996)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
           Economy, Trade and Industry of Japan; cDNA full insert sequencing:
           Research Association for Biotechnology (RAB); cDNA library
           construction: Helix Research Institute (HRI) (supported by Japan
           Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
           HRI, and Biotechnology Center, National Institute of Technology and
           Evaluation; clone selection for full insert sequencing: RAB and
           HRI.
FEATURES   Location/Qualifiers
    source   1..1796
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2NE2005317"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2NE2"
                     /note="cloning vector: pME18SFL3-mRNA from NT2 neuron
                     after the differentiation of NT2 neuronal precursor
                     cells."
BASE COUNT   490 a   365 c   463 g   478 t
ORIGIN

Query Match 28.6%; Score 885.4; DB 9; Length 1796;
Best Local Similarity 99.9%; Pred. No. 9e-190;
Matches 886; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   549 AGGCATAATGGTTATGAAAGCTTCTAGATGATGATGACGAATTCAGGATGGGAGCTCAGTAT 608
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   345 AGGCATAATGGTTATGAAAGCTTCTAGATGATGATGACGAATTCAGGATGGGAGCTCAGTAT 404

Qy   609 GCCAGAAAAAATGGAGAAAAAGCAATACAAACTGGGTGGACATTACCCAAGATTTTGAAGA 668
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   405 GCCAGAAAAAATGGAGAAAAAGCAATACAAACTGGGTGGACATTACCCAAGATTTTGAAGA 464

Qy   669 AGCTTGTGTCGAGATTAAAGTTGGGAGAAACTACTATTTCATGAGTACTATTTGGTCTTTTTGA 728
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   465 AGCTTGTGTCGAGATTAAAGTTGGGAGAAACTACTATTTCATGAGTACTATTTGGTCTTTTTGA 524

Qy   729 AGCCATGTCTCGTCTATTGAAATGATGGAATCCCAAGATGGGATTGATTGATTGGAAAACCA 788
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   525 AGCCATGTCTCGTCTATTGAAATGATGGAATCCCAAGATGGGATTGATTGATTGGAAAACCA 584

Qy   789 AGTTAATGAAAGTTCTCAATTTGAACAAGCTATCAAGAGGATGGCACTATTAAATTAA 848
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   585 AGTTAATGAAAGTTCTCAATTTGAACAAGCTATCAAGAGGATGGCACTATTAAAATTAA 644

Qy   849 AGATCTCACCTTGCCTGAACTGATGAGGATTAATGGATACACATGTTTTTGCTGTTTGATAAC 908
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   645 AGATCTCACCTTGCCTGAACTGATGAGGATTAATGGATACACATGTTTTTGCTGTTTGATAAC 704

Qy   909 GTGGTTAGAAGGCCATTCACTGGCACAGAGACAGTAGTTACGTTCCTTTACATTTCATTAATCC 968
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   705 GTGGTTAGAAGGCCATTCACTGGCACAGAGACAGTAGTTATTACGTTCCTTTACATTTCATTAATCC 764

Qy   969 AGACTTTTATAGAAGAATCTCCTATGAAAGGCTTTGCTCGGGAATCTTGAAAAATCTGTGA 1028
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   765 AGACTTTTATAGAAGAATCTCCTATGAAAGGCTTTGCTCGGGAATCTTGAAAAATCTGTGA 824
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   825 CATTGCAAGGGAAAAGTAAATAAAGCTGCTGTTTTTGAAGAGGAAGAGAATTTTCAGTCAAT 1088

Db   885 GACTTATGGATTTAAAATGGCTAACACAGTGTGACAGATCTTCGAGTTACAGGCATGCTAAA 944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1089 GACTTATGGATTTAAAATGGCTAACACAGTGTGACAGATCTTCGAGTTACAGGCATGCTAAA 1148

Db   945 AGATGTGGAGGATGACATCCAAAGAAGAGTAAAGAGTACTCGACAACAAGAGAGAGA 1004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1149 AGATGTGGAGGATGACATCCAAAGAAGAGTAAAGAGTACTCGACAACAAGAGAGAGA 1208

Db   1005 AAGAGGATCCAGAAGTTGAACTAGAAACACCAACAAATGTTAGCAGTATTCAGCGAGAGTGAA 1064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1209 AAGAGGATCCAGAAGTTGAACTAGAAACACCAACAAATGTTAGCAGTATTCAGCGAGAGTGAA 1268

Db   1065 ATTTACTCGTGTGTTACTGACAGTGCTTATAGCCTTTACTAGGAAGACCAGTGTGT 1124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1269 ATTTACTCGTGTGTTACTGACAGTGCTTATAGCCTTTACTAGGAAGACCAGTGTGT 1328

Db   1125 TGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTCTTCTGCCATTCATAATTC 1184
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1329 TGCAGAAGCTCAAAAATTGATGGTTCAAGCAGCAGATCTTCTTTCTTCTGCCATTCATAATTC 1388

Db   1185 ATTGCATCATTGGCATCCAGCCCCAGAATGATACTACAAAAGGAGGTC 1231
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy   1389 ATTGCATCATTGGCATCCAGCCCCAGAATGATACTACAAAAGGAGGTC 1435

RESULT 9
AX664135
LOCUS   AX664135   719 bp   DNA   linear   PAT 22-MAR-2003
DEFINITION Sequence 7 from Patent WO02057304.
ACCESSION   AX664135
VERSION   AX664135.1 GI:29164145
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Panzer,S.R., Lincoln,S.E., Altus,C.M., Dufour,G.E., Jackson,J.L.,
            Jones,A.L., Dam,T.C., Liu,T.F., Harris,B., Flores,V., Daffo,A.,
            Marwaha,R., Chen,A.J., Chang,S.C., Gerstin,E.H., Peralta,C.H.,
            David,M.H. and Lewis,S.A.
  TITLE   Secretory molecules
  JOURNAL   Patent: WO 02057304-A 7 25-JUL-2002;
            Incyte Genomics, Inc. (US)
FEATURES   Location/Qualifiers
    source   1..719
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: LI:462889.1:2001JAN12"
BASE COUNT   223 a   113 c   188 g   195 t
ORIGIN

Query Match 23.2%; Score 717.4; DB 6; Length 719;
Best Local Similarity 99.9%; Pred. No. 9.9e-152;
Matches 718; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   474 GGGGGCGGCCGAGCGCGGCGGCGTCGTTATTTCCCTGGGTCTCGGACACAGTGCGTGGCGCGGGT 533
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGGCGCGGCCGAGCGCGGCGGCGTCGTTATTTCCGTCGTTATTTCCGGACAGTGGCGTGGGCGGGGT 60

Qy   534 GACCACGGGAGGAGTAGGCCATAATGGTTATGAAAGCTTCGTAGATGATGAGATTCAGG 593
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GACCACGGGAGGAGTAGGCCATAATGGTTATGAAAGCTTCGTAGATGATGAGATTCAGG 120

Qy   594 ATGGGAGCTAGTAGTATGCCAGAAAAAAGTGGAGAAAAAATGCAATACAAAACTGGGTGGACATTAC 653

RESULT 10
AC102010

```
Db   121  |||||||||||||||||||||||||||||||||||||||||||||||||  180
Qy   121  ATGGGAGCTCAGTATGCCAGAGAAAATGGAGAAAAGCAATACAAACTGGGTGGACATTAC  180

Qy   654  CCAAGATTTGAAGAAGCTTGTCCAGAGAATTAAAACTTCATGATAAGCT  713
Db   654  |||||||||||||||||||||||||||||||||||||||||||||||||

Db   181  CCAAGATTTGAAGAAGCTTGTCGAGAATTAAAGTTGGGAGAACTACTTCATGATAAGCT  240
Qy   181  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   714  ATTTGGTCTTTTTGAACGCCATGTCGTCTATTGAAATGATGGATCCAAGAGATGGATGCTGG  773
Db   714  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   241  ATTTGGTCTTTTTGAAGCCCATGTCGTCTATTGAAAATGATGGATCCAAGAGATGGCTGG  300

Qy   774  CATGATTGGGAAACCAAGTTAATCCAATTTCAATTTTGAACAAGCTCATCATGATGGATGG  833
Db   774  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   301  CATGATTGGGAAACCAAGTTAATCAAGTTCTCAATTTTGAACAAGCTCATCAAGGATGG  360

Qy   834  CACTATTAAAATTAAAGATCTCACCTTGCCTGACATGATTAGGGGATTATTGGATTACAATCATGTTT  893
Db   834  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   361  CACTATTAAAATTAAAGATCTCACCTTGCCTGACATGATTAGGGGATTATTGGATTACAATCATGTTT  420

Qy   894  TTGCTGTTTGATAACGTGGTTAGAAGGCCCATTCACTGGCACAGACAGTATTTACGTTGCGT  953
Db   894  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   421  TTGCTGTTTGATAACGTGGTTAGAAGGCCATTCACTGGCACGACAGTATTTACGTTGCCT  480

Qy   954  TTACATTCATAATCCAGACTTTATAGAAGAGATCTTCGTATGTGAAGGCTTTTTGCTCTGGGAAT  1013
Db   954  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   481  TTACATTCATAATCCAGACTTTATAGAACATCCAAAGGCTTTTTCTGGGAAT  540

Qy   1014 CTTGAAAATCTGTGACATTGCAAAGGGAAAAAAGTAAAAAAGCTGCTGTTTTTGGGAAGGA  1073
Db   1014 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   541  CTTGAAAATCTGTGACATTGCAAGGGAAAAAAGTAAAATAAAAGCTGCGTGTTTTGAAGAGGA  600

Qy   1074 AGATTTCAGTCAATGACTTATGGATTAAAATGGCTAACAGTGTGACAGATCTTCGAGT  1133
Db   1074 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   601  AGATTTCAGTCAATGACTTATGGATTAAAATGGCTAACAGTGTGACAGATCTTCGAGT  660

Qy   1134 TACAGGCCATGCTAAAAGATGTGGAGGATGACATGCATGGAGTAAAGAGTAAAAGTACTCGAA  1192
Db   1134 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   661  TACAGGCCATGCTAAAAGATGTGGGGGATGACATGCCAAAAGAAGAGTAAAAGTACTCGAA  719
```

LOCUS   AC102010   147876 bp   DNA   linear   HTG 21-MAR-2003
DEFINITION Mus musculus clone RP24-531M24, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   AC102010
VERSION   AC102010.3 GI:29135693
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 147876)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE   Direct Submission
  JOURNAL   Unpublished

REFERENCE   2 (bases 1 to 147876)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE   3 (bases 1 to 147876)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Liu,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talanas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Mar 21, 2003 this sequence version replaced gi:28412092.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: 531_M_24
            Center clone name: L17873
            --------------- Summary Statistics
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 146099 bases at least Q40
            Consensus quality: 146869 bases at least Q30
            Consensus quality: 147004 bases at least Q20
            Insert size: 157000; agarose-fp
            Insert size: 147276; sum-of-contigs
            Quality coverage: 9.9 in Q20 bases; agarose-fp
            Quality coverage: 10.6 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces is
            * not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *   1   84958: contig of 84958 bp in length
            *   84959   85058: gap of 100 bp
            *   85059   87411: contig of 2353 bp in length
            *   87412   87511: gap of 100 bp
            *   87512   89933: contig of 2422 bp in length
            *   89934   90033: gap of 100 bp
            *   90034   94076: contig of 4043 bp in length
            *   94077   94176: gap of 100 bp
```

```
              *   94177   102536: contig of 8360 bp in length
              *  102537   102636: gap of 100 bp
              *  102637   118954: contig of 16318 bp in length
              *  118955   119054: gap of 100 bp
              *  119055   147876: contig of 28822 bp in length

FEATURES             Location/Qualifiers
     source          1..147876
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-531M24"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature    1..84958
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    85059..87411
                     /note="assembly_fragment"
     misc_feature    87512..89933
                     /note="assembly_fragment"
     misc_feature    90034..94076
                     /note="assembly_fragment"
     misc_feature    94177..102536
                     /note="assembly_fragment"
     misc_feature    102637..118954
                     /note="assembly_fragment"
     misc_feature    119055..147876
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT      46111 a    27663 c    27986 g    45516 t    600 others
ORIGIN

Query Match     21.5%;  Score 664.2;  DB 2;  Length 147876;
Best Local Similarity   72.6%;   Pred. No. 2.6e-139;
Matches 1105;  Conservative   0;  Mismatches 293;  Indels 125;  Gaps  14;

Qy   1647 TGTTCTTCAAGATCTCTGTTACTAAACCAC-TTTCCTGGTGGATAACAAAAACGGTCTTTG 1705
          |||||||| | |||||||||||||||||||| |||||||| ||||||||||  || |||||
Db  72297 TGTTCTTTCAAGATCTCTGTTACAAAACCACTTTTCCTCGTGGGTAACACAACATCTTTTG 72356

Qy   1706 G-AACTCATCTCATGCAAGACATGGTGAAAGAATGCACTCGGTCTCTTTGTCAGATCCTCC 1764
          | ||||| | |||||| |||||  ||||||| ||  | | |||||||||||  |||| |||
Db  72357 GAAACTCCCTCATGAAAGACATGGTGAAATATTACCTTTGGTCCTTCATT---AGTCCTC 72413

Qy   1765 GAGTGCTTTCCCCAAGTGCTACCCTATATAATAATAATCACCAGGGTAAGGACTGTATCGACT 1824
          |||| ||||    || ||| ||||||| ||||||||  |||||  |||||||| || |||
Db  72414 CAGTGCTATCCCCTAAGTGTTGCCTATATAATGACACCAGGCTAAGGACTGTATCGACT 72473

Qy   1825 CCTTTGTTACTCACTGTGTTCGGCCAATTCTCGTTAGTCTTCTTATTCAGATCCATCCATGGACATAAACA 1884
          |||||||||||||||||||||||||||| ||| || ||| ||||||||||||||||||||||||||||
Db  72474 ACTTTGTTACTCACTGTGTTTGGGCCAATTTGTTGTAGTCTTCAGAGTCATAGCCATAACA 72533

Qy   1885 GGGCTCGACAGAGAGAATAAGCTTGGTGTCATAATTCTGAGGAATTGCCACCTTGCAGGATG 1944
          ||| |  || ||  ||| |||| ||||||| || ||||  || |||| ||||||| |||||
Db  72534 GGGCTCAGCAGAGAGTCAAACTTGGGTCACATCCTTGAGAAGTTTGTTACCTTGCAAGATG 72593

Qy   1945 AGTTTATGAC--------------ATTTTATTTTAATAGCGCAGAGA 1976
          |||||| |||              | |||| || ||||| |||||||
Db  72594 AGTTTTATACTCCAGAGGATGAAGTTCTTGAAGAAATGTAACACTTTGCAGACA 72653

Qy   1977 AGGTTGATCGCAGCCGCTTCACACCATGCTGTTTGAAACAAGGAACCCCAAAGGCAACATTTGG 2036
          |||||||| || | |||||||||||| ||||||||||| | ||||  || ||||||||||||||
Db  72654 ACTTTGATCCAGTGCTTCACACTCCGTTGTTGTGAAAACAGGAGCCGTGAGAGACTACATCTGA 72713

Qy   2037 ACTTTTAGGTACCTGGGTCCTTACCATAACCTTGCCATTATGATATACTACCTTCTAA 2096
          ||||||||| | |||||||| |||||||| |||||||||| || | ||||  ||| ||
Db  72654 ACTTTGATCCAGTGCTTCACACTCCGTTGTTGTGAAAACAGGAGCCGTGAGAGACTACATCTGA 72713

Qy   2037 CCTGTTTAGGTACCTGGGTCCTTACCATAACCTTGCCATTATGATATACTACCTTCTAA 2096
          |||||||||| |||||||||| |||||||| |||||||||| || | ||||  || ||
Db  72714 CCTGTTTAGGAATCTGAGTTCTTTACCAAAATCTTCAAATTATGATTCAGTATCTCGCCCA 72773

Qy   2097 GTGGCTTTGAATTGGAAACTTCACACGAGTATGCACGAGAGATATGCACGAGAGATC 2156
          |||||||| ||||||||||||||||| |||||| || || |||| | | | || | |
Db  72774 TTGCCTTTGAGCTGGAGCTGTACACGCATGAGAGTCTTACTCCATGTACTGGTACCTCT 72833
```

              ... (alignment continues in right column) ...

```
Qy   2157 CTGGAATTCCTTTAGCCATGGTTGATGTCAACCATTGATGTCGTGCCGATGGCTCAAAATGG 2216
          ||||||| | || | |||| || |||||||  ||||| | |||| || ||| |  || || |
Db  72834 CTGAGTTCCTGTATGCGTCGCTGCTGATGTCAA-------------ACTTGACAGCTCTCAGATGG 72883

Qy   2217 CAGAGGAAAGGATAATGGAGAGCAGCAGAAAAGCGGCCGAGAAAGGCCGTAGTAG------- 2259
          | |||| |||| ||||||||| ||| || | || |  | ||||| |||  |||||
Db  72884 CAGAGGTGAAGATAATAGGAGAGCACGCAGAAAGGACACAGCAGCAAAAACAAAAATAACAA 72943

Qy   2260 --------------TAAAAAAAACAAAAAAAAAAAAAAAGAAAAAGTTCGCCCCATTGAGCCGAGAGATC 2308
                        |||||||||||||||||||||||||||||||||||||||||||||
Db  72944 ACAAAAAAAAAAAAAAAACAAAAAAGAAAAAGAAAAGAAAGTTCGTTGAGCCCAAGAGATC 73003

Qy   2309 ACAATGAGCCAAGACATATCAGAACATGTGTCCTGGAATGTTTAAAACCATGGTAGCATTT 2368
          ||||| |||| |||| |||| |||| | |||| ||||| |||| |||| ||||| |||||
Db  73004 ATAATCAGCCAGCCTATCAGAA-ATGTGGCGTGGAATGTTCAGAATGATGGTCATGTTT 73062

Qy   2369 GACATGGACCGGCAAAGTACGTAAACCGAAGTTGAGCTTGAGTAGTGAACAAGTTCGGTAT 2428
          ||||||| | ||||||| | ||| |||||||||||||||||||||||||| ||| |||
Db  73063 GACATGGATGGCAAAGTGCTCAAGCCCAAGTTTGAGCTTGAGTGATCAGGTCTCGGTAT 73122

Qy   2429 GAACACAAGTTTTGCTCCATTCAAACACGTGTGATG------------ACCCGCGCCAGTGCACT 2480
          || |||||||||  ||||| ||||||| |||  |             |||| | || ||||
Db  73123 GAGCACAGGTCTGCGTCTTTGTTCAAACAGTGTGATGTTTACAACTGCCAGAGAACT 73182

Qy   2481 ACTTACAGTTCAAGGAAATGTCTGACCTCAATAAAATAAATATAGTGTT-----ATATTC 2540
          | |||||||||||| |||| | || ||||||||||  | ||||||||||     ||||||
Db  73183 ACTTGCCGTTCAAGGAAAATGGCTGACCTCAATAAAATAATATAGCCCCTCCTCTCAGGTCTCCTG 73242

Qy   2541 AACTGTATGTGCAGCTAGTAAGCACTTTCAACAGCACAGGCAAAAATGATATTATTGGAAAATATTC 2600
          |||| || || || |  |  | ||||| |||| |||| |||||||||| |||| ||| |||||||
Db  73243 AGCTTTATGTGGCAGCTAATAAGCACTTTCAACAGGCAAAAATGATCCTTGAAATGTCC 73302

Qy   2601 CTAACCCGGACCATAGGAGGTTAATAGAATTTTAAAGGTTGCCAAACCCAACTTTGTGGTTA 2660
          || |||| | |  | | ||| |||||| |||| |||| ||  || |||||||| || |||
Db  73303 CCAATCCAGCGAGTGAGGTCAGTAGATAATCCTTAAGGTTGTTAAGTCCAACTTTGTGCTTA 73362

Qy   2661 TGAAGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCTCCTGAATTGAATTTCT 2720
          ||||||| ||| ||||| || ||| ||| |||| | || ||||||| || ||||  |||
Db  73363 TGAAGCTCTTGGCAGGAGGACACAAAAGAGAGTCAAAGGTTCCTCCCGAAATTTGGACTTCT 73422

Qy   2721 CTGCTCATAATATATTTTCCTGTTGTGTAAATGAATGA------------AAGTGCAAAGAACGACTGT 2780
          |||||||| ||| ||||||| || |||| ||| ||||             ||| || || ||
Db  73423 CTGTTCACAAAATACTTTCATATTGTGAAAGTGAATTGAAAAGAACGACTGT 73482

Qy   2781 AAAGGGGCAGAGTCTTCTTTCAGAGACCCAACTCTTAGAGGGGCACATCACCAGGCTCCACAT 2840
          | ||| |||| || | ||| || || | |||||||| ||||| || ||||||||| ||||
Db  73483 GT----------CCACTCTAGGGGATACATCAGTCTGCCACAT 73515

Qy   2841 CACGGGAAGTGAGATGGAGTTTCTTGGGT---AACAACTCATTATAAGGAATACTTTT-A 2895
          |||||| || | || ||| ||| |||||    ||| | | | ||| || ||||| ||| |
Db  73516 ATGGTGAGGTAAAGCGGATTTCTTGAATGACAGAGCAGCTCCTTAAAAAGAATACTTTTCA 73575

Qy   2896 GTTTGACAGCCTTATATGACAATGAAATGAAAACTGCTGTTTTAAAGTGGTTTATTAGTTC 2955
          |||| || |||| || ||| || || || | | |||||||| || || |||||||||||
Db  73576 GTTTAACACTTTACCTGGCATGAGATAAACTGCTCTTTTAAAGTGGTTTATTATTATG--- 73632

Qy   2956 CATGGAAGAAACTGGTCTTATTAGAATGCAATGAGAAACGTTATATATGGTTTTATTACAGAT 3015
          ||||||||||| ||| | || || |||| |||| ||||  ||| |||||||| | |||||
Db  73633 ------------GGTGTAACTGGGATAGAAGTATTTGGTGAAATAGTTACATGATTTATTATAC 73685

Qy   3016 TTAAT--CACAAATCATTTTTATGAATGAATGAATGGTGAAAATAGTTGTTTAAAAGGTTAA 3073
          ||||   |||| | || ||| | || ||| | || |||  ||| | ||| || || || ||
Db  73686 TTAATTGTTTAAAACATTTTAAGGAATGATCATTGAGTG-AAAATGGTGTTTGTAAAGATTAA 73744

Qy   3074 TAAAATTTCTTGACAAAAAAAAAAAA 3096
          ||||| ||| || |||||||||  ||
Db  73745 TAAATTTCTTGACAAAAAATTAA 73767
```

RESULT 11
AL353743
LOCUS   AL353743   181864 bp   DNA   linear   PRI 08-SEP-2001
DEFINITION Human DNA sequence from clone RP11-213G2 on chromosome 9, complete

sequence.

ACCESSION   AL353743
VERSION   AL353743.22 GI:15552931
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 181864)
  AUTHORS   Corby,N.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT   On Sep 11, 2001 this sequence version replaced gi:1426784.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-213G2 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-213G2 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-213G2 is at 181864 in this
            sequence. The true right end of clone RP11-65C15 is at 111320 in
            this sequence. The true right end of clone RP11-202I11 is at 100 in
            this sequence.
FEATURES   Location/Qualifiers
     source   1..181864
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-213G2"
                     /clone_lib="RPCI-11.1"
     misc_feature   39727
                     /note="sequence from overlapping clone bA65C15 (AL161453)."
     misc_feature   11957..116075
                     /note="sequence from overlapping clone bA65C15 (AL161453).
                     Assembly confirmed by restriction digest."
     unsure   11488..117564
                     /note="sequence from overlapping clone bA65C15 (AL161453).
                     Assembly confirmed by restriction digest."
BASE COUNT   53151 a 36268 c 36336 g 56109 t
ORIGIN

Qy   1 TTCCTCACGAAACTCCCACGGCGGCTGTATAGGAAACATAAATCCGTTGTCAGGCAGCAGTA   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174760 TTCCTCACGAAACTCCCACGGCGGCTGTATAGGAAACATAAATCCGTTGTCAGGCAGCAGTA 174819

Query Match   17.8%; Score 551.4; DB 9; Length 181864;
Best Local Similarity   99.8%;
Matches 552; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 TTCCTCACGAAACTCCCACGGCGGCTGTATAGGAAACATAAATCCGTTGTCAGGCAGCAGTA   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174760 TTCCTCACGAAACTCCCACGGCGGCTGTATAGGAAACATAAATCCGTTGTCAGGCAGCAGTA 174819

Qy   61 GCACGCGGTTGCTCTCGGAGCTTGGCTGCTGTTCGTTGTTGCTGCTGCAACCACTAAGGTCTACG   120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174820 GCACGCGGTTGCTCTCGGAGCTTGGCTGCTGTTCGTTGTTGCTGCTGCAACCACTAAGGTCTACG 174879

Qy   121 CAAACCTCCACGGTTTCCTTCCGTCCTTCGCGTCCACCTTTCTAAGAAATTCCCAGAGGGCA   180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174880 CAAACCTCCACGGTTTCCTTCCGTCCTTCGCGTCCACCTTTCTAAGAAATTCCCAGAGGGCA 174939

Qy   181 GCGCGAGACGGGGCGGGGGTCTGAGACTCGGGGCTTGGGCTGTCGTTCGTTCGGGAAACCGCCGCCACTA   240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174940 GCGCGAGACGGGGCGGGGGGTCTGAGACTCGGGGCTTGGGCTGTCGTTCGTTCGGGAAACCGCCGCCACTA 174999

Qy   241 CCCAGGAGACTCCGACCAGAGGGGTGAAAAAGATAACTTCCGGGTCTTCGCCATGTCTCTAATC   300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 175000 CCCAGGAGACTCCGACCAGAGGGGTGAAAAAAGATAACTTCCGGGTCTTCGCCATGTCTCTAATC 175059

Qy   301 TCGCGGAGAGAGGAAGGCGGGCGCCCGCCATCGCCGACGAGGAGGAGCGGGGGGGTG   360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 175060 TCGCGGAGAGAGGAAGGCGGGCGCCCGCCATCGCCGACGAGGAGGAGCGGGGGGGTG 175119

Qy   361 TGGCCCGGGGAGGCGGGAAGTCCCCGGGAGTAAGGGAGGAGGAGGGAGGTCGCGCGTCCCGG   420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 175120 TGGCCCGGGGAGGCGGGAAGTCCCCGGGAGTAAGGGAGGAGGAGGGAGGTCGCGCGTCCCGG 175179

Qy   421 GCATACGCCATGCGTGCACGCGCTGCGGTCGGGGGGGGCTGAGAGGGAGGGGGGGGGGGGG   480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 175180 GCATACGCCATGCGTGCACGCGCTGCGGTCGGGGGGGGCTGAGAGGGAGGGGGGGGGGGGG 175239

Qy   481 GCCGAGGCGGGCGTCGTTATTTCCGTGGTCCGGACAGTGTCGGGTGGGGGGGGTGACCACG   540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 175240 GCCGAGGCGGGCGTCGTTATTTCCGTGGTCCGGACAGTGTCGGGTGGGGGGGGTGACCACG 175299

Qy   541 GGAGAAGTAGGCA   553
         |||||||||||||
Db 175300 GGAGAAGTAGGTA 175312

RESULT 12
AL161447
AL161447

LOCUS   AL161447   160115 bp   DNA   linear   PRI 16-MAR-2002
DEFINITION Human DNA sequence from clone RP11-379P1 on chromosome 9, complete
            sequence.
ACCESSION   AL161447
VERSION   AL161447.20 GI:19572746
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 160115)
  AUTHORS   Johnson,C.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. Clone requests: clonerequest@sanger.ac.uk
            humquery@sanger.ac.uk E-mail enquiries: clonerequest@sanger.ac.uk
COMMENT   On Mar 21, 2002 this sequence version replaced gi:15787718.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP

Left column

```
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-379P1 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
This sequence is the entire insert of clone RP11-379P1 The true
left end of clone RP11-280P22 is at 137145 in this sequence. The
true right end of clone RP11-65C15 is at 38497 in this sequence.
FEATURES             Location/Qualifiers
     source          1..160115
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-379P1"
                     /clone_lib="RPCI-11.2"
     misc_feature    complement(32505..32575)
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
     misc_feature    74372..74537
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data."
     misc_feature    74706..74920
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data."
     misc_feature    complement(74706..74920)
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
     misc_feature    74921
                     /note="Tandem repeat. Forced join. Gap size estimated to
                     be approximately 300bp by restriction digest data."
     misc_feature    129108
                     /note="Tandem repeat. Forced join. Gap size estimated to
                     be approximately 200bp by restriction digest data."
BASE COUNT    46921 a  33388 c  34629 g  45177 t
ORIGIN

Query Match     12.9%;  Score 398.4;  DB 9;  Length 160115;
Best Local Similarity   99.8%;  Pred. No. 5.2e-79;
Matches  399;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy     2697 AAGTTCCTCCCTGRAATTTGATTTCTCTGCTCATAAATATTTTTCCTGTTGTGAAACTTGTTT 2756
            ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    43136 AGGTTCCTCCCTGRAATTTGATTTCTCTGCTCATAAATATTTTTCCTGTTGTGAAACTTGTTT 43195

Qy     2757 GAGAGAGAGACTGGGGAGGAGGTGGCCATAAAAGGGGCAGAGAGTCTTCTTTTCAGACCCAACTCCTTAG 2816
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    43196 GAGAGAGAGACTGGGGAGGAGGTGGCCATAAAAGGGGCAGAGAGTCTTCTTTTCAGACCCAACTCCTTAG 43255

Qy     2817 AGGGCACATCACCAGGCTCCACATCACCGGGAGAAGTGAGAGATGGGATTTCTTGGGGTAACAACTC 2876
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    43256 AGGGCACATCACCAGGCTCCACATCACCGGGAAGTGAGAGATGAATTTCTTGGGGTAACAACTC 43315

Qy     2877 AATTATAAGGAATACTTTTAGTTTGACAGCCTTATATCACATGAATGAAAAACTCCTGTGTTTT 2936
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    43316 AATTATAAGGAATACTTTTAGTTTGACAGCCTTATATGACATGAATGAAAAACTGCTGTGTTTT 43375

Qy     2937 AAAGTGGTTTATTTATTGTTCCATGGGAAGAAAACTGGTCTTATTGAAATGCATTGATTGAAACGTT 2996
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    43376 AAAGTGGTTTATTTATTGTTCCATGGGAAGAAAACTGGTCTTATTCATGGAATGCATTGATTGAAACGTT 43435

Qy     2997 ATATGGTTTATTTACAGATTTAATCACAAATCATTTTTTATGAATGATTGAGTGAAAATA 3056
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    43436 ATATGGTTTATTTACAGATTTAATCACAAATCATTTTTTATGAATGATTGAGTGAAAATA 43495

Qy     3057 GTGTTTATAAAGGTTAATAAANTTCCTGACAAAAAAAAAAAA 3096
            ||||||||||||||||||||| ||||||||||||||||||||
Db    43496 GTGTTTATAAAGGTTAATAAANTTCCTGACAAAAAAAAAA 43535
```

Right column

```
RESULT 13
HSA334821/c
LOCUS       HSA334821                446 bp    DNA     linear   PRI 18-JUL-2002
DEFINITION  Homo sapiens genomic sequence surrounding NotI site, clone
            NB6-015R.
ACCESSION   AJ334821
VERSION     AJ334821.1  GI:15879239
KEYWORDS    
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
            Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
            Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashuba,V.I.,
            Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
  TITLE     NotI flanking sequences: a tool for gene discovery and verification
            of the human genome
  JOURNAL   Nucleic Acids Res. 30 (14), 3163-3170 (2002)
  MEDLINE   22135767
   PUBMED   12136098
REFERENCE   2  (bases 1 to 446)
  AUTHORS   Zabarovsky,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
            Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
            Sweden
FEATURES             Location/Qualifiers
     source          1..446
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="NB6-015R"
BASE COUNT      111 a    102 c    120 g    113 t
ORIGIN

Query Match     10.4%;  Score 321.4;  DB 9;  Length 446;
Best Local Similarity   99.7%;  Pred. No. 6e-62;
Matches  322;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy       1 TTCCTCACGAAAACTCCCAGGCGCTGTATAGGAGAAACATAAATCCGTTGTCAGGCAGCAGTA 60
           ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     323 TTCCTCACGACACTCCCAGGCGCTGTATAGGAGAAACATAAATCCGTTGTCAGGCAGCAGTA 264

Qy      61 GCACGCTGTTGCTCTTCGGAGACGTTGGCTGCTTCGGTCGGTTCGTTCGTCGCAACCCACTAAGGTTCTACG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     263 GCACGCTGTTGCTCTTCGGAGACGTTGGCTGCTTCGGTCGGTTCGTTCGTCGCAACCCACTAAGGTTCTACG 204

Qy     121 CAAAACCTCCACGGTTTCCTTCGCGTTCGCCTTGCGGTCACCCTTTCTAAGAAAATTCCCAGAGGGCA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     203 CAAAACCTCCACGGTTTCCTTCGCGTTCGCCTTGCGGTCACCCTTTCTAAGAAAATTCCCAGAGGGCA 144

Qy     181 GCGCAGGACGGGCGGGGCGGCCTCTGAGACTTCGCGGCGCTCCGCGCCTCTTTCCGGGGAACCGCGCCCACTA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     143 GCGCAGGACGGGCGGGGCGGCCTCTGAGACTTCGCGGCGCTCCGCGCCTCTTTCCGGGGAACCGCGCCCACTA 84

Qy     241 CCCAGGGACTCCGACAGGCGGTGAAAAAGAGATAACTTCCGGGGTCTCGCGGATCGTCTCTAATC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83 CCCAGGGACTCCGACAGAGGGTGAAAAAGAGATAACTTCCGGGGTCTCGCGGATCGTCTCTAATC 24

Qy     301 TCGGCGGAGAAGGAAGGCGGCCGC 323
           |||||||||||||||||||||||
Db      23 TCGGCGGAGAAGGAAGGCGGCCGC 1

RESULT 14
HSA334413/c
LOCUS       HSA334413                743 bp    DNA     linear   PRI 18-JUL-2002
DEFINITION  Homo sapiens genomic sequence surrounding NotI site, clone
            NR1-ND24R.
```

ACCESSION   AJ334413
VERSION   AJ334413.1 GI:15878831
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 743)
  AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
             Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
             Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashuba,V.I.,
             Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
  TITLE   NotI flanking sequences: a tool for gene discovery and verification
             of the human genome
  JOURNAL   Nucleic Acids Res. 30 (14), 3163-3170 (2002)
  MEDLINE   22131767
   PUBMED   12136098
REFERENCE   2 (bases 1 to 743)
  AUTHORS   Zabarovsky,E.R.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
             Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
             Sweden
FEATURES   Location/Qualifiers
     source   1..743
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="NR1-ND24R"
BASE COUNT   219 a   154 c   166 g   203 t   1 others
ORIGIN

Query Match   9.7%; Score 299; DB 9; Length 743;
Best Local Similarity   95.4%; Pred. No. 7.9e-57;
Matches   308; Conservative   0; Mismatches   15; Indels   0; Gaps   0;

Qy   1 TTCCTCACGAAAACTCCCACAGGCGTCGGCTGTTGTATAGGAAAACATAAAATCCGTTGTTGTCACACGAGTA 60
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||| ||||||||||||||
Db   323 TTCCCTCACGAAAACTCCCAGGCGCTGTATAGAAAAACATAAATCCGTTGTCACACGAGTA 264

Qy   61 GCACGCTGTTGCTCTCTGGGAGCTTGGCTGCTGCTCGTCGTTCGTTCGTTCGTTCGCAACCACTAAGGTCTACG 120
              ||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   263 GCACGCTGTTGCTCTCCGAAGCTTGGCTGCTGTTCGTTCGTTCGTCGCAACCACTAAGGTCTACG 204

Qy   121 CAAAACCTCACGGTTTCCTTCGCCTTCGGCTCACCTTTTTTAAGAAAATTCCCAGAGGGCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   203 CAAAACCTCACGGTTTCCTTCGCCTTCGGCTCACCTTTTTTAAGAAAATTCCCAGAGGGCA 144

Qy   181 GCGCAGACCGGGCGGGGGCCTCGAGACTCCCGGGCGTCCGGGGGAACCGCCCACTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   143 GCGCAGACCGGGCGGGGGGCTCGAGACTCCGGGCGTCCGGGGGAACGGCCCACTA 84

Qy   241 CCCAGGACTCCGACACAGAGGGTGAAAAAAGATAACTTCCGGTCTCGCGATCGTCTCTTAATC 300
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   83 CTCAGGACTCCGTCAGAGAAAAAAACCGATTACTTCCGGTCTCGCGATCGTCTCTTAATC 24

Qy   301 TCGCCGAGAAGAAGAAGGCGGGGCCGC 323
              |||||||||||||||||||||||||||
Db   23 TCGCCGAGAAGAAGAAGGCGGGGCCGC 1

RESULT 15
LOCUS   HSA342328   642 bp   DNA   linear   PRI 18-JUL-2002
DEFINITION Homo sapiens genomic sequence surrounding NotI site, clone
            NR1-ND24C.
ACCESSION   AJ342328
VERSION   AJ342328.1 GI:15886745
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 642)
  AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
            Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
            Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashuba,V.I.,
            Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
  TITLE   NotI flanking sequences: a tool for gene discovery and verification
            of the human genome
  JOURNAL   Nucleic Acids Res. 30 (14), 3163-3170 (2002)
  MEDLINE   22131767
   PUBMED   12136098
REFERENCE   2 (bases 1 to 642)
  AUTHORS   Zabarovsky,E.R.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
            Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
            Sweden
FEATURES   Location/Qualifiers
     source   1..642
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="NR1-ND24C"
BASE COUNT   73 a   209 c   265 g   93 t   2 others
ORIGIN

Query Match   7.6%; Score 236.4; DB 9; Length 642;
Best Local Similarity   99.6%; Pred. No. 1.2e-42;
Matches   237; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   316 GCGGCCGCCATCGGCCGAACGGAGCGGAGCGGAGCGGTGCGAGGAGGAGGGGTGTTGCCGGAGGCGCG 375
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCGGCCGCCCATCGGCCGAACGGAGCGGAGCGGTGCGAGGAGGAGGGGTGTTGCCGGAGGCGCG 60

Qy   376 AAGTCCCCGGGAGTAAGGGAGGAGGGGGGGGCGGGGGGGGGCGGGGGGGGGGCGGTGCGCGTCCCGGGCCATACGGATCATGCCGTG 435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGTCCCCGGGAGTAAGGGAGGAGGGGGGGGGGGGGGGGCGGGGGGGGGGCGGTCCCGGGCCATACGGATCATGCCGTG 120

Qy   436 CACGCTGCCGGTGGGGCTGGGCGTGGGCTGAGAGGGAGGGGAGGGGGGGGCGGGGGGCGTCG 495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CACGCTGCCGGTGGGGCTGGGCGTGGGCTGAGAGGGAGGGGAGGGGGGGGCGGGGGGCGTCG 180

Qy   496 TTATTTCCGTGGGTCCGGACAGTGGCGTGGGCGCGGGGGTGACCACGGGAGGAAGTAGGCA 553
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   181 TTATTTCCGTGGTCCGGACAGTGCGTGGGCGCGGGGGTGACCACGGGAGGAAGTAGGTA 238

Search completed: November 25, 2003, 01:06:53
Job time : 11149 secs

OM nucleic - protein search, using frame_plus_n2p model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: November 25, 2003, 03:13:51 ; Search time 106 Seconds
(without alignments)
10683.169 Million cell updates/sec

Title: US-10-001-857-42
Perfect score: 5439
Sequence: 1 ttcctcacgaaactcccagg..........atttcttgacaaaaaaaa 3096

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 667484 seqs, 182883703 residues
Total number of hits satisfying chosen parameters: 1334968

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10001857/runat_24112003_162856_9623/app_query.fasta_1.3271
-DB=Published_Applications_AA -QFMT=fastan SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10001857 @CGN_1_1_149 @runat_24112003_162856_9623
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

SUMMARIES

Result No.	Score	%	Query Match	Length	DB	ID	Description
1	3796	69.8	733	14		US-10-001-857-145	Sequence 145, App
2	523	3.0	523	12		US-10-017-161-1982	Sequence 1982, Ap
3	146.5		574	12		US-10-168-097A-76	Sequence 76, Appl
4	146.5		574	12		US-10-239-431A-38	Sequence 38, Appl
5	145	2.7	559	12		US-09-858-155A-26	Sequence 26, Appli
6	145	2.6	559	10		US-09-168-097A-26	Sequence 26, Appli
7	145	2.6	559	12		US-10-239-431A-33	Sequence 33, Appl
8	145	2.6	559	12		US-10-116-370-2	Sequence 2, Appli
9	144	2.6	1134	14		US-10-001-873-50	Sequence 50, Appl
10	142.5	2.6	800	12		US-10-029-386-32198	Sequence 32198, A
11	142	2.6	721	15		US-10-086-464-5	Sequence 5, Appl
12	142	2.6	5877	15		US-10-142-515-11	Sequence 11, Appl
13	141.5	2.6	5935	15		US-10-243-243A-8	Sequence 8, Appl
14	140		802	9		US-09-823-246-2	Sequence 2, Appli
15	139.5	2.5	1243	15		US-09-196-935A-4	Sequence 4, Appli
16	138.5	2.5	641	11		US-09-959-987-2	Sequence 2, Appli
17	138.5	2.5	647	9		US-09-925-299-1002	Sequence 1002, Ap
18	138.5	2.5	647	15		US-09-925-299-1002	Sequence 1002, Ap
19	136.5	2.5	647	14		US-10-086-464-2	Sequence 2, Appli
20	136.5	2.5	647	14		US-10-086-464-4	Sequence 4, Appli
21	136.5	2.4	520	12		US-10-168-097A-36	Sequence 36, Appli
22	134.5	2.4	520	12		US-10-239-431A-34	Sequence 34, Appl
23	134	2.4	351	12		US-10-239-431A-22	Sequence 22, Appl
24	134	2.4	542	10		US-10-239-431A-10	Sequence 10, Appl
25	133.5	2.4	507	14		US-10-078-547-24	Sequence 24, Appl
26	132.5	2.4	206	15		US-10-156-761-14830	Sequence 14830, A
27	132.5	2.4	797	15		US-10-156-761-10907	Sequence 10907, A
28	132	2.4	503	14		US-10-078-547-2	Sequence 2, Appli
29	131	2.4	528	12		US-09-840-746-20	Sequence 20, Appli
30	131	2.4	538	10		US-09-976-740-43	Sequence 43, Appli
31	131	2.4	538	14		US-10-023-529-43	Sequence 43, Appli
32	131	2.4	538	14		US-10-023-523-43	Sequence 43, Appli
33	130	2.4	235	15		US-10-050-704-104	Sequence 104, Appl
34	130	2.4	455	12		US-10-032-585-7399	Sequence 7399, Ap
35	129.5	2.4	501	12		US-10-168-097A-46	Sequence 46, Appl
36	129.5	2.4	501	12		US-10-239-431A-35	Sequence 35, Appl
37	129	2.4	377	15		US-10-149-819-7	Sequence 7, Appl
38	129	2.3	505	12		US-10-168-097A-56	Sequence 56, Appl
39	129	2.3	505	12		US-10-239-431A-36	Sequence 36, Appl
40	126.5	2.3	405	12		US-10-017-161-2256	Sequence 2256, App
41	126.5	2.3	1046	15		US-10-156-761-10088	Sequence 10088, A
42	126	2.3	416	15		US-10-043-487-282	Sequence 282, App
43	126	2.3	505	12		US-10-168-097A-12	Sequence 12, App
44	126	2.3	505	12		US-10-239-431A-32	Sequence 32, Appl
45	125	2.3	550	10		US-09-976-740-47	Sequence 47, Appl

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
US-10-001-857-145

Sequence 145, Application US/10001857
Publication No. US20020183500A1
GENERAL INFORMATION:
APPLICANT: Macina, Roberto
APPLICANT: Recipon, Herve
APPLICANT: Chen, Sei-Yu
APPLICANT: Sun, Yongming
APPLICANT: Liu, Chenghua
TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P
FILE REFERENCE: DEX-0273
CURRENT APPLICATION NUMBER: US/10/001,857
CURRENT FILING DATE: 2001-11-20
PRIOR APPLICATION NUMBER: 60/252,054
PRIOR FILING DATE: 2000-11-20
NUMBER OF SEQ ID NOS: 208
SOFTWARE: PatentIn version 3.1
SEQ ID NO 145
LENGTH: 733
TYPE: PRT
ORGANISM: Homo sapien
US-10-001-857-145

```
Alignment Scores:
Pred. No.:              5e-294
Score:                  3796.00
Percent Similarity:     99.59%
Best Local Similarity:  99.59%
Query Match:            69.79%
DB:                     69.79%

                        Length:        733
                        Matches:       732
                        Conservative:  0
                        Mismatches:    0
                        Indels:        1
                        Gaps:          0

US-10-001-857-42 (1-3096) x US-10-001-857-145 (1-733)
```

```
Qy   556  ATGGTTATGAAAGCTTCTGTAGATGATGACGATTCAGGATGGGAGCTCAGTATGCCAGAA  615
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  MetValMetLysAlaSerValAspAspAspAspSerGlyTrpGluLeuSerMetProGlu   20

Qy   616  AAAAATGGAGAAAAACAAATACAAACTGGGTGGACATTACCCAGAGATTTGAAGAGCTTGT  675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    21  LysAsnGlyGluLysGlnThrAsnTrpValAspIleThrGlnAspPheGluGluAlaCys   40

Qy   676  CGAGAATTAAAGTTGGGAGAACTACTTCATGATGAAGCTATTTGGTCTTTTTGAAGCCATG  735
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    41  ArgGluLeuLysLeuGlyGluLeuLeuHisAspLysLeuPheGlyLeuPheGluAlaMet   60

Qy   736  TCTGCTATTGAAATCATGGAACCCCAAGATGGAGTGGATGGAAACCAAGTTAAT  795
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  SerAlaIleGluIleMetGluProLysMetAspProLysMetGlyAsnGlnValAsn   80

Qy   796  CGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGATGGCACTAATAAAATTAAAGATCTC  855
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    81  ArgLysValLeuAsnPheGluGlnAlaIleLysAspGlyThrIleLysIleLysAspLeu  100

Qy   856  ACCTTGCCTGAACTGATAGGGGAATTATGGGATACATGTTTTTGCTGTTTGATAACGTGGTTA  915
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101  ThrLeuProGluLeuIleGlyIleMetAspThrCysPheCysCysLeuIleThrTrpLeu  120

Qy   916  GAAGGCCATTCACTGGCACACAGTATTACGTGCCTTTACAATCCATAAATCCAGACTTT  975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  GluGlyHisSerLeuAlaGlnThrValPheThrCysLeuTyrIleHisAsnProAspPhe  140

Qy   976  ATAGAAGAATCCTGCTATGAAGGCTTTTTGCTCTCGGAAATCCTGTGACATTGCA  1035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141  IleGluGluSerProAlaMetLysAlaPheLeuAlaLeuGlyIleLeuLeuLysCysAspIleAla  160

Qy  1036  AGGGAAAAAGTAAATAAAGCTGCTGTTTTGGAAGAGGAAGAAGATTTTCAGTCAAGTGAACTTAT  1095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   161  ArgGluLysValAsnLysAlaAlaValAlaValPheGluGluGluGluAspPheGlnSerMetThrTyr  180

Qy  1096  GGATTTAAAATGGCTAACAGTTGACAGATCTTCGAGTTACAGCGCATGTGTAAAAGATGTG  1155
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  GlyPheLysMetAlaAsnSerValThrAspLeuArgValThrArgIleTyrMetLeuLysAspVal  200

Qy  1156  GAGGATGACATGCAAAGAGACTAAGAGTACTCGACGAGGAGAAGAAGAGAT  1215
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201  GluAspAspMetGlnArgArgValLysSerThrArgSerThrArgGlnGlyGluGluArgAsp  220

Qy  1216  CCAGAAGTTGAACTAGAACAACAAATGTTTAGCCAGTAATGAGAGTGAAAATTTACT  1275
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   221  ProGluValGluLeuGluGlnHisGlnCysLeuAlaValPheSerArgValLysPheThr  240

Qy  1276  CGTGTGTTACTGACACTGCTTATAGCCTTTTACTAAGAAAGACCAGTGCTGTTGTTCCAGAA  1335
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  ArgValLeuLeuThrValLeuIleAlaPheThrLysLysGluThrSerAlaValAlaGlu  260

Qy  1336  GCTCAAAAATTGAATGGGTTCAAGCAGCAGGATCTTCTTCTTCTGCCATTCAATAATTCATTGCAT  1395
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261  AlaGlnLysLeuMetValGlnAlaLaAaAspLeuLeuSerArgAlaIleHisAsnSerLeuHis  280

Qy  1396  CATGGCATCCAGGCGCCCAGAATGATACTAGCAAAAGAGAATCATCCAATTATGATGGGTTTT  1455
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   281  HisGlyIleGlnAlaGlnAsnAspThrThrLysGlyAspHisProIleMetMetGlyPhe  300

Qy  1456  GAACCCCTTTGAACCAGAGGGTTACTTCCCACCTTCCCTCGATATGCAAAAATTAATT  1515
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  GluProLeuValAsnGlnArgLeuLeuProThrPheProArgTyrAlaLysIleIle  320
```

```
Qy  1516  AAAAGGGAAGAAAATGGTGAAGTAGTATTTTGCCAAGATTAAATAAGAATAAAAACTGTCTGT  1575
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321  LysArgGluGluAsnGlyGluValValPheTyrPheAlaArgLysAsnAsnLys  340

Qy  1576  GAGGTTGTGAATTTAACAAATTTACATTGTATCCTGGATTTTTCTGTGAATTAGTGAA  1635
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   341  GluValValAsnLeuThrAsnLeuHisCysIleLeuAspPhePheCysGluPheSerGlu  360

Qy  1636  CAGTCACCAATGTGTTCTTTCAAGATCTCTGTTCAAAACCACTTTCCTCGTGGATAACAA  1695
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  GlnSerProCysValLeuSerArgSerLeuLeuGlnThrThrPheLeuValAspAsnLys  380

Qy  1696  AAGGTCTTTGGAAACTCATCCTGCCAAGACACATGGTGAAAGAGATGCACTTCGGTCTTTTGTC  1755
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   381  LysValPheGlyThrHisLeuMetGlnAspMetValLysAspAlaLeuArgSerPheVal  400

Qy  1756  AGATCCTCCGAGTGCTTTCCCCCAAGTGCTTACCCTATAATAATAATCACCAGCGCTAAGACT  1815
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   401  -AspProPro-ValLeuSerProLysCysTyrLeuTyrAsnAsnHisGlnHisLysAspC  420

Qy  1816  GTATCGACTCCTTGTTCACTGTTGAAACAGGAACCCAAAGGCCATTCGTAGTCCCTTATTCAGATCCATG  1875
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  ysIleAspSerProPheValThrHisCysSerValArgProPheCysSerLeuIleLeuIleGlnIleHisG  440

Qy  1876  GACATAACAGGGCTCGACAGAGAGTAAGCTTGGTCATATTCTTGAGGAATTTGCCACCT  1935
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   441  lyHisAsnArgAlaArgGlnArgAspSerLysLeuGlyHisIleLeuLeuGluGluPheAlaThr  460

Qy  1936  TGCAGGATGAGTTATGACATTTTATTTTAAATAGGCA-GAGAAGGTTGATGCGAGCGCTTC  1994
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   461  euGlnAspGlyPheMetThrPheTyrPheAsnArgAlaGluLysValAspAlaAlaLeuH  480

Qy  1995  ACACCATGCTGTTGAAACAGGAACCCAAAGGCAACATTGGCCTGTTTAGGTACCTGGG  2054
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  ysIleAspSerSerPheValThrHisCysSerValArgProPheCysSerLeuIleLeuIleHisG  500

Qy  2055  TCCTTTACCATAACCTTCGCATTATGATACAGTACCTTTCTAAGTGCTTTGAATTGGAAC  2114
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   501  alLeuTyrHisAsnArgIleMetIleGlnTyrLeuSerGlyPheGluLeuLeuGluL  520

Qy  2115  TCTACAGTATGCACGAGGACTACTAATTACATATAATTGGTATGTCTCTGAATTCCTTTACGGAT  2174
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   521  luGlyHisGluGlyTyrTyrIleTyrTrpTyrValSerGluPheLeuTyrAlaT  540

Qy  2175  GGTTGATGTCAACATTGAGTCGTCGTCCCGATGGCCTCTCAAATGCCAGAGGAAAGGATAATGG  2234
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  rpLeuMetSerThrLeuSerArgArgAlaAspGlySerGlnMetAlaGluGluArgIleMetG  560

Qy  2235  AAAGAGCAGCAGAAAAGGCCGTAGTTTAAAAAAAAACAAAAGAAAAAAAAAGTTCGCCCAT  2294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   561  luGluGlnGlnLysGlyArgSerSerLysThrLysSerLysSerLysSerValArgProL  580

Qy  2295  TGAGCCCGAGAGAGATCACAATGAGCCAAGCGAATCACGAGAACATGTGTGGCTGGAAGTTTAAAA  2354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   581  euSerArgGluIleThrMetSerGlnAlaTyrGlnAsnMetCysAlaGlyMetPheLysT  600

Qy  2355  CCATGGTAGCCATTTGACATGGACCGGCAAGTACGTAAACCCGAAGTTTGAGCTGATAGTG  2414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  hrMetValAlaIleAspMetAspArgGlnLysTyrValAsnProLysPheGluLeuIleVal  620

Qy  2415  AACACAGTTCGGTATGACACAGTTTGTCCATTCAACACTGTGATGAACCCGCCGCGCAG  2474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   621  LysGlnValArgTyrGluHisSerArgPheAspAsnSerValMetThrProProProValV  640

Qy  2475  TGCACTACTTACAGTTCAGGGAAAGTCTGACCTCAATAATAATAATAGCCCTCCTCCTCAGT  2534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   641  alHisTyrLeuGlnPheArgGluLysSerAspLeuAsnLysTyrSerProProProGlnS  660

Qy  2535  CTCCTGAACTGTATGTGGCAGCTAAGCACTTTCAACAGGCAAAAAAATGATATTGGAA  2594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  erProGlyLeuTyrValAlaAlaSerLysHisPheGlnGlnAlaLysMetIleLeuGluA  680
```

```
Qy    2595 ATATTCCTAACCCGACCATGAGGTTAATAGAAATTTTAAAGGTTGCAAACCAACTTTG 2654
           ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     680 snileProAsnProAspHisGluValAsnArgIleLeuLysValAlaLysProAsnPheV 700

Qy    2655 TGGTTATGAAGTTAATGGCACGAGGACACAAAAAGGAATCTAAAGTTCCTCCTGAATTTG 2714
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     700 alValMetLysLeuLeueuAlaGlyGlyHisLysGluSerLysSerValProGluGluPheA 720

Qy    2715 ATTTCTCTGCTCATAAATAATATTTCCTCGTTGTGAAACTTGTT 2755
           ||||||||||||||||||||||||||||||||||||
Db     720 spPheSerAlaHisLysTyrPheProValValLysLeuVal 733
```

RESULT 2
US-10-017-161-1982
; Sequence 1982, Application US/10017161
; Publication No. US20030143568A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 1982
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (259)..(265)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (373)..(376)
; OTHER INFORMATION: Variable amino acid
US-10-017-161-1982

Alignment Scores:
Pred. No.:
Score: 0.000233
Percent Similarity: 165.00
Best Local Similarity: 38.32%
Query Match: 20.11%
DB: 3.00%
                        12

US-10-001-857-42 (1-3096) x US-10-017-161-1982 (1-523)

```
Qy    1337 GCTTCTGCAACAGCACTGGTCTCTTCTTAGTAAAGGCTATAAGCACTGTCAGTAACACA 1278
           ||||||||||| |||||||  |||   ||   |
Db       3 AlaThrAlaThrThrIle-----------------------ThrThrThrThr 14

Qy    1277 CGAGTAAATTCACTCGTCGAATACTGCTAAACATTGTTGGTGTTTCTRAGTTCAACTTCT 1218
           |||||||    |||||
Db      15 ThrIleIleIleIleIleThrAlaThrThrIleIleIleThrThrIleThr 34

Qy    1217 GGATCTCTTTCTTCTCCTTGTCGACTTCGAGTACTTCTTTACTCTTCTTTGCATGTCATCC 1158
           :::
Db      35 ThrIleThrIleThrThrThrIleThrAlaThrThrIleIleIleIleThrAla 54

Qy    1157 TCCACATCTTTTAGCATGCCTGTAACTCGGAGATCTGTCACACACTGTTAGCCATTTAAAT 1098
           |||||::|:||  |||
Db      55 ThrThrThrIleIleIleThrThrThrThrThrIleThrThrIleThrIle------ 72

Qy    1097 CCATAAGTCATTGACTGAAAATCTTCCTCTTCAAAAAACAGCAGCTTTATTTACTTTTTCC 1038
           ||||||||::|  ||||
Db      73 -----------------ThrIleThrAlaThrThrIleThrIleIle 83
```
```

RESULT 3
US-10-168-097A-76
; Sequence 76, Application US/10168097A
; Publication No. US20030166245A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL OF THE RECHERCHE SCIENTIFIQUE
; APPLICANT: INSTITUT CURIE
; TITLE OF INVENTION: WASP FAMILY PROTEIN FRAGMENTS, AND USES THEREOF
; FILE REFERENCE: IFB99WASP
; CURRENT APPLICATION NUMBER: US/10/168,097A
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 76
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe

```
Qy    1037 CTTGCAAGTCACAGGATTTTCAAGATTTTCAAGATTCCAGAGATTCCAGAGCAAAAGCCTTCATAGCAGGATCTTCT 978
           :::::  :||:
Db      84 IleIleIleThrAlaThrThrIleThrIleIleThrIleThrThrThrThr 103

Qy     977 ATAAAGTCTGGATTATGAAATGTAAAGGCACGTAAATACTGTCTGTGCCAGTGAATGGCCT 918
           |||  |:::
Db     104 IleThr-----------------IleThrThrIleThrAlaThr 112

Qy     917 TCTAACCACGTTATCAAACAGCAAAAACAATGTATCCATAATCCCATCAGTTCAGGCAAG 858
           ::||
Db     113 ThrIleThrThrValIleThrThrIleThrIleIleThrIleThrIleThr 132

Qy     857 GTGAGAATCTTTAAATTTTAATAGTAGCCATCCTTGATAGCTTGTTGTTCAAAATTGAGAACTTT 798
           :|||||    ||||
Db     133 ThrThrThrThrIleThrIleIleIleSerSerTyrLeuIleThrThrThrThrThr 152

Qy     797 CGATTAACTTGGTTTCCAAATCATGCCAGCATCCATCTTGGGAATCCATCATTCAATAGCA 738
           |||  |||  :::
Db     153 ThrAlaThrThrThrIleThrThrThrIleThrIleAsnThrIleThrIleIle 172

Qy     737 GACATGGCTTCAAAAAAGACCAAATAGCTTATCATGAAGTAGTTCTCCAACTTTAATTCT 678
           |||:::   ||:::
Db     173 AsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn-------AsnHisSerAsnAsn:yrIleThr 189

Qy     677 CGACAAGCTTCTTCAAAATCTTGGGGTAATGTCCACCCAGTTGTGTTGTATTGGCTTTTCTCCATT 618
           |||||
Db     190 ThrThrThrThrThrThrAsnThrAsnIleThrThrProThrThr-------ThrThr 206

Qy     617 TTTTCTTCGGCATACTGAGCTCCCATCCTGAATCGTCATCATCATCTACAGAAGCTTTCATATACC 558
           ::|    ||||
Db     207 ThrThrThrIleThrThrThr-ThrIleAsnAsnHisHis-TyrHisHisHisHisH 226

Qy     557 ATTATGCCTACTT-CTCCCGTGGTCACCCGGCGCCGCCACCGACTGTTCCGGACCACGGAA 499
           |:::   |||  ||||| |||      ||||
Db     226 isHisGlnHisHisThrProProProThrThrThrThrThrThrT 246

Qy     498 TAACGACG-------------------------------- 491
           ||||||
Db     246 hrThrThrThrAsnThrThrAsnAsnAsnAsnAsnAsn*******P 266

Qy     490 -------------------------------CGGCCTCGGCGCCGGCGCCC 454
           :::   ||||||||||||
Db     246 roProProProProProProGlnProProProProProProProPro- 325

Qy     393 CCTTACTCCCGGGACTTCGGCGCCTCCCGGCCCACACCCCCTCCGCCCTCCGT 334
           ||||  |||||||||||||
Db     306 roProProProProProProProProProProProProProProPro- 306

Qy     333 TCGGCGCGATGGCGCGGCGCCCT 314
           ||||||||
Db     326 ---------------ProPro 327
```

US-10-168-097A-76

Alignment Scores:
Pred. No.:
Score: 0.00721
Percent Similarity: 146.50   Matches: 53
Best Local Similarity: 39.41%   Conservative: 14
Query Match: 31.18%   Mismatches: 64
                        2.67%   Indels: 39
                        12   Gaps: 8

US-10-001-857-42 (1-3096) × US-10-168-097A-76 (1-574)

```
Qy  544  CTCCCGTGGTCACCCGCGCCGCCGCCCAGCCACTGTCCGGACCACGGAAATAACGACGCGCCGCCT  485
         ||||||  ||  |   ||| |||:::|
Db  336  LeuProProProProProProArgSerAsnAlaAlaGlySerIleProLeuProPro  355

Qy  484  CGG-------------------CCGCGCGGCGCGCCCCCCTCCCCTCTCAGCCCGACCGGGCAGCGGT  437
         :::                   ||||||||||||||||||| |||
Db  356  GlnGlyArgSerAlaProProProProProProArgSerAla---ProSerThrGlyArg  374

Qy  436  GCACGCCATGCGTATGCCCGGGACGCGGC-----------------GACCCCGCCCCTCTCCCTTA  389
         |||                                          ||||||||||||| |||
Db  375  GlnProProProLeuSerSerSerArgAlaValSerAsnProProAlaProProProAla  394

Qy  388  CTCCCGGGGACTTGGCGTCCCCGGGCCACACCC------------------------------  356
         |||||||||   |||::|||||||||||
Db  395  IleProGlyArgSer---AlaProAlaLeuProProLeuGlyAsnAlaSerArgThrSer  413

Qy  355  ----------------------CCTCCCTCGCCACCGCCTCCCGTTCGGCCGA  326
                               ||||||
Db  414  ThrProProValProThrProProSerLeuProProSerAlaProProSerLeuProPro  433

Qy  325  TGGCGCGCCGCCTTCTTCTTCGCGGATTAGAGACGATCGCGAGACGGAAGTTATCTTT  266
         |||||||||                                   :::
Db  434  SerAlaProProSerLeu-----ProMetGlyAlaProAlaAlaProAlaProLeuProPro  451

Qy  265  TTTCACCCTCTGTGCGGAGTCCTGGGGTAGTGGGC----------------GGTTCC  227
         |||||||||                          |||
Db  452  SerAlaProIleAlaProLeuProAlaGlyMetProAlaAlaProAlaProLeuProPro  471

Qy  226  CGGAAAGAGGCGGAGCGCCCGAGTCTTCAGAGCCCCGTCTGGGGCTGCGCCCTCTGGGAAT  167
         |||||||||||                    |||||:::|||
Db  472  AlaAlaProAlaProProProAlaProAlaAlaAla---ProValAlaSer  490

Qy  166  TTCTTAGAAAGGTGACGGACGGAAGGCGGAAGG  137
         |||    :::||| |||||
Db  491  IleAlaGluLeuProGlnGlnAspGlyArg  500
```

RESULT 4
US-10-239-431A-38
Sequence 38, Application US/10239431A
Publication No. US20030170726A1
GENERAL INFORMATION:
APPLICANT: FRADELIZE, JULIE
APPLICANT: FRIEDERICH, EVELYNE
APPLICANT: GOLSTEYN, ROY M.
APPLICANT: LOUVARD, DANIEL
APPLICANT: NOIREAUX, VINCENT
APPLICANT: SYKES, CECILE
TITLE OF INVENTION: PEPTIDE SEQUENCES COMPRISING ONE OR MORE UNITS BINDING
TITLE OF INVENTION: TO PROTEINS OF THE Ena/VASP FAMILY, AND THEIR USES
FILE REFERENCE: 0508-1032
CURRENT APPLICATION NUMBER: US/10/239,431A
CURRENT FILING DATE: 2002-09-23
PRIOR APPLICATION NUMBER: PCT/FR01/00843
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: FR 00/03637
PRIOR FILING DATE: 2000-03-22
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 38
LENGTH: 574
TYPE: PRT

; ORGANISM: Schizosaccharomyces pombe
US-10-239-431A-38

Alignment Scores:
Pred. No.:
Score: 0.00721
Percent Similarity: 146.50   Matches: 53
Best Local Similarity: 39.41%   Conservative: 14
Query Match: 31.18%   Mismatches: 64
                        2.67%   Indels: 39
                        12   Gaps: 8

US-10-001-857-42 (1-3096) × US-10-239-431A-38 (1-574)

```
Qy  544  CTCCCGTGGTCACCCGCGCCGCCGCCCAGCCACTGTCCGGACCACGGAAATAACGACGCGCCGCCT  485
         ||||||  ||  |   ||| |||:::|
Db  336  LeuProProProProProProArgSerAsnAlaAlaGlySerIleProLeuProPro  355

Qy  484  CGG-------------------CCGCGCGGCGCGCCCCCCTCCCCTCTCAGCCCGACCGGGCAGCGGT  437
         :::                   ||||||||||||||||||| |||
Db  356  GlnGlyArgSerAlaProProProProProProArgSerAla---ProSerThrGlyArg  374

Qy  436  GCACGCCATGCGTATGCCCGGGACGCGGC-----------------GACCCCGCCCCTCTCCCTTA  389
         |||                                          ||||||||||||| |||
Db  375  GlnProProProLeuSerSerSerArgAlaValSerAsnProProAlaProProProAla  394

Qy  388  CTCCCGGGGACTTGGCGTCCCCGGGCCACACCC------------------------------  356
         |||||||||   |||::|||||||||||
Db  395  IleProGlyArgSer---AlaProAlaLeuProProLeuGlyAsnAlaSerArgThrSer  413

Qy  355  ----------------------CCTCCCTCGCCACCGCCTCCCGTTCGGCCGA  326
                               ||||||
Db  414  ThrProProValProThrProProSerLeuProProSerAlaProProSerLeuProPro  433

Qy  325  TGGCGCGCCGCCTTCTTCTTCGCGGATTAGAGACGATCGCGAGACGGAAGTTATCTTT  266
         |||||||||                                   :::
Db  434  SerAlaProProSerLeu-----ProMetGlyAlaProAlaAlaProAlaProLeuProPro  451

Qy  265  TTTCACCCTCTGTGCGGAGTCCTGGGGTAGTGGGC----------------GGTTCC  227
         |||||||||                          |||
Db  452  SerAlaProIleAlaProLeuProAlaGlyMetProAlaAlaProAlaProLeuProPro  471

Qy  226  CGGAAAGAGGCGGAGCGCCCGAGTCTTCAGAGCCCCGTCTGGGGCTGCGCCCTCTGGGAAT  167
         |||||||||||                    |||||:::|||
Db  472  AlaAlaProAlaProProProAlaProAlaAlaAla---ProValAlaSer  490

Qy  166  TTCTTAGAAAGGTGACGGACGGAAGGCGGAAGG  137
         |||    :::||| |||||
Db  491  IleAlaGluLeuProGlnGlnAspGlyArg  500
```

RESULT 5
US-09-858-155A-2
Sequence 2, Application US/09858155A
Patent No. US20020137049A1
GENERAL INFORMATION:
APPLICANT: Mark, Robert
APPLICANT: Young, Kathleen H.
APPLICANT: Wood, Andrew
TITLE OF INVENTION: PABLO, A POLYPEPTIDE THAT INTERACTS WITH BCL-XL, AND
TITLE OF INVENTION: USES RELATED THERETO
FILE REFERENCE: GNN-005
CURRENT APPLICATION NUMBER: US/09/858,155A
CURRENT FILING DATE: 2001-05-15
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 559
TYPE: PRT
ORGANISM: Homo sapiens
US-09-858-155A-2

Alignment Scores:
Pred. No.:
Score: 0.00942
                        145.00   Matches: 48
                                  Length: 559
```

Left column

```
Percent Similarity:        40.52%              Conservative:  14
Best Local Similarity:     31.37%              Mismatches:    61
Query Match:               2.64%               Indels:        30
DB:                        10                  Gaps:          8

US-10-001-857-42 (1-3096) x US-09-858-155A-2 (1-559)

Qy  559 CCATTATGCCTA-----------------------CTTCTCCCGTGGTCACCCGGCCGGCCACGGCACT 515
        ||||                              | ||| |||||
Db  294 ProThrCysIleSerSerAlaThrGlyLeuIleGluAsnArgProGlnSerProAlaThr 313

Qy  514 GTCCGGACCACGGAAATAACGACGCCCGCCCGCCGCCGCCCCTCCCCTCTCAGCC 455
        |||||
Db  314 GlyArgThrProValPheValSerProThrProProProProProProProLeuProSer 333

Qy  454 CAGCCCGACCGGCAGCCGTGCACGCATGCGTATGCCCGGGACGCGGGACCCC-----GCC 401
        ||||
Db  334 AlaLeuSerThrSerSerLeuArgAlaSerMetThrSerThrProProProProProValPro 353

Qy  400 CCCTCTCCCTTACTCCCGGGCCACACCCCCT------ 353
        |||
Db  354 ProProProProProAlaThrAlaLeuGlnAlaAlaProAlaValProProProAla 373

Qy  352 ----------------------CCCTCGCCACCG--------CCTCCGTTC 332
        :::|           |||||   |||
Db  374 ProLeuGlnIleIleAlaProGlyValLeuHisProAlaProProProIleAlaProProLeu 393

Qy  331 GGCCGGATGGCGGCGCGCCTTCTCTTCTCCGGAGATTAGAGACGATCCCGAGACCGGAAGTT 272
        :::                  ::::         |||
Db  394 ValGlnProSerProPro-----ValAlaArgAlaAlaProValCys-----GluThr 409

Qy  271 ATCTTTTTCACCCTCTGTCGGAGTCCTGGGCGGTTCCCGAAAGAGGCGGAG 212
        :::
Db  410 ValProValHisProLeuProGlyGlnGlyValGlnGlyLeuProProProProPro 429

Qy  211 CCCGGAGTCTCAGAGCCCGCCCCGTCTCGCGTGCCTCT 173
        |||||
Db  430 ProProLeu-----------ProProProGlyIleArgProSer 440


RESULT 6
US-10-168-097A-26
; Sequence 26, Application US/10168097A
; Publication No. US20030166245A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL OF THE RECHERCHE SCIENTIFIQUE
; APPLICANT: INSTITUT CURIE
; TITLE OF INVENTION: WASP FAMILY PROTEIN FRAGMENTS, AND USES THEREOF
; FILE REFERENCE: IFB99WASP
; CURRENT APPLICATION NUMBER: US/10/168,097A
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 26
; LENGTH: 559
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-168-097A-26

Alignment Scores:
Pred. No.:                 0.00942
Score:                     145.00             Length:        559
Percent Similarity:        40.52%             Matches:       48
Best Local Similarity:     31.37%             Conservative:  14
Query Match:               2.64%              Mismatches:    61
DB:                        12                 Indels:        30
                                              Gaps:          8

US-10-001-857-42 (1-3096) x US-10-168-097A-26 (1-559)

Qy  559 CCATTATGCCTA-----------------------CTTCTCCCGTGGTCACCCGGCCGGCCACGGCACT 515
        ||||                              | ||| |||||
Db  294 ProThrCysIleSerSerAlaThrGlyLeuIleGluAsnArgProGlnSerProAlaThr 313

Qy  514 GTCCGGACCACGGAAATAACGACGCCCGCCCGCCGCCGCCCCTCCCCTCTCAGCC 455
```

Right column

```
Db  314 GlyArgThrProValPheValSerProThrProProProProProProProLeuProSer 333

Qy  454 CAGCCCGACCGGCAGCCGTGCACGCATGCGTATGCCCGGGACGCGGGACCCC-----GCC 401
        ||||
Db  334 AlaLeuSerThrSerSerLeuArgAlaSerMetThrSerThrProProProProProValPro 353

Qy  400 CCCTCTCCCTTACTCCCGGGCCACACCCCCT------ 353
        |||
Db  354 ProProProProProAlaThrAlaLeuGlnAlaAlaProAlaValProProProAla 373

Qy  352 ----------------------CCCTCGCCACCG--------CCTCCGTTC 332
        :::|           |||||   |||
Db  374 ProLeuGlnIleIleAlaProGlyValLeuHisProAlaProProProIleAlaProProLeu 393

Qy  331 GGCCGGATGGCGGCGCGCCTTCTCTTCTCCGGAGATTAGAGACGATCCCGAGACCGGAAGTT 272
        :::                  ::::         |||
Db  394 ValGlnProSerProPro-----ValAlaArgAlaAlaProValCys-----GluThr 409

Qy  271 ATCTTTTTCACCCTCTGTCGGAGTCCTGGGCGGTTCCCGAAAGAGGCGGAG 212
        :::
Db  410 ValProValHisProLeuProGlyGlnGlyValGlnGlyLeuProProProProPro 429

Qy  211 CCCGGAGTCTCAGAGCCCGCCCCGTCTCGCGTGCCTCT 173
        |||||
Db  430 ProProLeu-----------ProProProGlyIleArgProSer 440


RESULT 7
US-10-239-431A-33
; Sequence 33, Application US/10239431A
; Publication No. US20030170726A1
; GENERAL INFORMATION:
; APPLICANT: FRADELIZE, JULIE
; APPLICANT: FRIEDERICH, EVELYNE
; APPLICANT: GOLSTEYN, ROY M.
; APPLICANT: LOUVARD, DANIEL
; APPLICANT: NOIREAUX, VINCENT
; APPLICANT: SYKES, CECILE
; TITLE OF INVENTION: PEPTIDE SEQUENCES COMPRISING ONE OR MORE UNITS BINDING
; TITLE OF INVENTION: TO PROTEINS OF THE Ena/VASP FAMILY, AND THEIR USES
; FILE REFERENCE: 0508-1032
; CURRENT APPLICATION NUMBER: US/10/239,431A
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: PCT/FR01/00843
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: FR 00/03637
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 33
; LENGTH: 559
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-239-431A-33

Alignment Scores:
Pred. No.:                 0.00942
Score:                     145.00             Length:        559
Percent Similarity:        40.52%             Matches:       48
Best Local Similarity:     31.37%             Conservative:  14
Query Match:               2.64%              Mismatches:    61
DB:                        12                 Indels:        30
                                              Gaps:          8

US-10-001-857-42 (1-3096) x US-10-239-431A-33 (1-559)

Qy  559 CCATTATGCCTA-----------------------CTTCTCCCGTGGTCACCCGGCCGGCCACGGCACT 515
        ||||                              | ||| |||||
Db  294 ProThrCysIleSerSerAlaThrGlyLeuIleGluAsnArgProGlnSerProAlaThr 313

Qy  514 GTCCGGACCACGGAAATAACGACGCCCGCCCGCCGCCGCCCCTCCCCTCTCAGCC 455
        |||||
Db  314 GlyArgThrProValPheValSerProThrProProProProProProProLeuProSer 333
```

```
Qy  454 CAGCCGACCGGCAGCGTGCACGGCATGGTATGCCCGGATGGGCCGGGGACCCC------GCC 401
        |||
Db  334 AlaLeuSerThrSerSerLeuArgAlaSerMetThrSerThrProProProValPro      353

Qy  400 CCCTCTCCCTTACTCCCGGGGACTTCG--CGCTCCCCGGGCCGCCCACCCCCCT------    353
        :::       |||
Db  354 ProProProProProAlaThrAlaLeuGlnAlaLeuProProProProAla            373

Qy  271 ATCTTTTTCACCCTCGTCGGAGTCCTGGGTAGTGGGCGGTTCCCGGAAAAGAGGCGGAG    212
        |||||||
Db  410 ValProValHisProLeuProGlnGlyLeuGlnGlyLeuProProProProProPro      429

Qy  211 CCCGGAGTCTCAGAGCCCCGCCCCCCGTCTGCGCCGCTGCCCTCT                  173
        :::
Db  430 ProProLeu------ProProProGlyIleArgProSer                        440

Qy  352 ------CCTCCGTTC                                                332
Db  394 ValGlnProSerProPro------ValAlaArgAlaAlaAlaProValCys------GluThr 409

Qy  331 GGCCGGATGGGCGGCGCCGGCCTTCTTCTTCGCCGGAGGATTAGAGACGATCGGCGGAGACCGGAAGTT 272
```

RESULT 8
US-10-116-370-2
```
; Sequence 2, Application US/10116370
; Publication No. US20030190709A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: PABLO, A POLYPEPTIDE THAT INTERACTS WITH BCL-XL, AND
; TITLE OF INVENTION: USES RELATED THERETO
; FILE REFERENCE: AM100012-D2
; CURRENT APPLICATION NUMBER: US/10/116,370
; CURRENT FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.0
;
; SEQ ID NO 2
; LENGTH: 559
; TYPE: PRT
; ORGANISM: Homo sapiens
```

US-10-116-370-2

Alignment Scores:
```
Pred. No.:
Score:                    0.00942
                          145.00
Percent Similarity:       40.52%        Length:          559
Best Local Similarity:    31.37%        Matches:         48
Query Match:              2.64%         Conservative:    14
                                        Mismatches:      61
                                        Indels:          30
                                        Gaps:            8
```

US-10-001-857-42 (1-3096) x US-10-116-370-2 (1-559)

```
Qy  559 CCATTATGCCTA------CTTCTCCCGTGGTCACCGCGCCCAGCGCACT             515
        |||||                 |||                  :::
Db  ... ProThrCysIleSerSerAlaArgThrGlyLeuIleGluIleAsnArgProGlnSerProAlaThr 313
```

RESULT 9
US-10-001-873-50
```
; Sequence 50, Application US/10001873
; Publication No. US20020160388A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; APPLICANT: Turner, Leah
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P...
; FILE REFERENCE: DEX-0275
; CURRENT APPLICATION NUMBER: US/10/001,873
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,055
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,496
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patentin version 3.1
;
; SEQ ID NO 50
; LENGTH: 1134
; TYPE: PRT
; ORGANISM: Homo sapien
```

US-10-001-873-50

Alignment Scores:
```
Pred. No.:
Score:                    0.0145
                          144.00
Percent Similarity:       37.97%        Length:          1134
Best Local Similarity:    33.54%        Matches:         53
Query Match:              2.62%         Conservative:    7
                                        Mismatches:      64
                                        Indels:          34
                                        Gaps:            8
```

US-10-001-857-42 (1-3096) x US-10-001-873-50 (1-1134)

```
Qy  559 CCATTATGCCTACTTCTCCCGTGGTCACCGCGCCACTGTCCGGACACGGAA            500
        |||
Db  976 ProLeuProLeuArgLeuProLeuProProProPro---                        988

Qy  499 ATAACGACGCGCGCCTGCGCCGCGCCGCCCCCTCA-----GCCAGCCGAC            446
        |||                  |||
Db  989 ---LeuProArgProHisProProProProLeuProLeuProProPro             1007

Qy  445 CGGCAGCGTGCA------CGCATGGCTATGCCCGGAGCGGGCGGACCCCGCC          401
        |||||
Db 1008 GlnThrArgThrLeuProAlaAlaArgThrMetArgGlnPro------ProPro       1023

Qy  400 CCCTCTCCTTACTCCCGGGGGACTTGGCGCTCCCCGGGCCACACCCCCTCGCCA--      344
        |||                        |||
Db 1024 ProArgLeuAlaLeuProArgArgArgSerProProArgProSerArgProAla       1043

Qy  343 ------CCGCTCCGTTGGGCGCGATGGCGGGCGCCGCGCTTCTCTTCTCGGGAGA       299
                   |||                  |||
Db 1044 ArgArgGlyProArgProThrProGlnAlaArgArgProArgProArgProArgArg    1063

Qy  298 TTAGAGACGATCGGCGGAGACCGGAAGTTATCTTTTTTCACCCTCGTGCGGAGTCCTGGGTA 239
        |||
Db 1064 Leu------LeuArgSerPro------HisSerLeuCysSerProArgLeu          1076
```

```
Qy    238 GTGGGGCGGTTCCCGGAAAGAG------GCGGAGCCCGGAGTCCTTCAGAGAGCCCCGTCT 185
              |||||                    |||  ||||||||||||||||||||||||||||
Db   1077 ArgProGlyProArgGlyAlaAspProArgArgGluArgAlaSerThrSerProProArg 1096

Qy    184 GCGCTGCCCTCTGGGAATTTCTTAGAAAGGTGACGCGAAGGCGGAAGGAAACCG 131
              |||                      |||                    |||
Db   1097 SerTrpProSerGlySerAlaCysArgProTrpProArgThrGlyProArgSerPro 1114
```

RESULT 10
US-10-029-386-32198

; Sequence 32198, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32198
; LENGTH: 800
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007663.28
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3

US-10-001-857-42 (1-3096) x US-10-029-386-32198 (1-800)

Alignment Scores:
Pred. No.:   0.0169
Score: 142.50
Percent Similarity: 35.66%
Best Local Similarity: 17.72%
Query Match: 2.59%

Length: 800
Matches: 76
Conservative: 77
Mismatches: 202
Indels: 74
Gaps: 11

```
Db   1334 TCTGCAACAGCACTGGTCTCTTCTTAGTAAGGCTATAAGCACTGTTCACAACACGA 1275
             ||||: ||||:||||:|||||||||||||||||||||||||:|||:|||:||
Qy    290 ThrThrIleThrIleThrThrIleThrThrIleThrIleThrThrThrThr 309

Qy   1274 GTAAATTC------ACTCTGCTGAATACTGTAAACATTGTGGTGTTCTAGTTCA 1224
             |||            ::        ||: ||||||||||||||||
Db    310 IleThrIleThrThrThrThrThrIleThrThrThrThrThrThr 329

Db   1223 ACTTCTCGGATCTCTTCTTCTCTTCTTGTCGAGTACTCTTTACTCTTCTTTGCATG 1164
             ||||:||                  |:|||||||||||||||||||||:||
Qy    330 ThrThrThrIleThrThrIleThrThrIleThrThrThrThrThrIle 349

Qy   1163 TCATCCTCCACATCTTTTAGCATGCCTGTAACTCGGAAGATCTGTCACACTGTTAGCCATT 1104
             ||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    350 ThrThrThrThrThrIleThrThrThrThrThrIleThrIleThrThrThrThr 369

Db    350 ThrThrThrThrThrIleThrThrThrThrThrIleThrIleThrThrThrThr 369

Qy   1103 TTAAATCCATAAGTCATTGAACTAGAAAAATCTTCCTCTTCAAAAACACGAGCTTTATTTACT 1044
             ||||||||||||||||||||||||||||||||||||||||||||||||:||
Db    370 IleThrThr------ThrThrThrIleThrThrIleThrThrIleThrIleThr 384

Qy   1043 TTTTCCCTTGCAATGTCACAGATTTCAAGATTCCAGAGCAAAAGCCTTCATPAGCAGGA 984
             :::                              ||||             |||
Db    385 ThrThrThrThrThrThrIleThrThrIleThrThrThrIleThrThrIleThrThr 404

Qy    983 TCTTCTATAAAGTCTGGATTATGAAATGTAAAGGCACGTAAATACTGTCTGTGCCAGTGAA 924
             :::::|||                                      ::: :::
Db    405 ThrThrIleThrThrThrIle------------ThrThrThrThrThrIleThrThrThr 419
```

RESULT 11
US-10-086-464-5

; Sequence 5, Application US/10086464
; Publication No. US20020199218A1
; GENERAL INFORMATION:
; APPLICANT: GORING, Daphne R. et al.
; TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES
; FILE REFERENCE: P 25,762-A USA
; CURRENT APPLICATION NUMBER: US/10/086,464
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 10/069,304

```
Db    405 ThrThrIleThrThrThrIle--------------ThrThrThrThrThrIleThr 419
             :::
Qy    923 TGGCCTTCTAACCACGTTATCAAACACGAAAAAACATGTATCCATCAGTTCA 864
             |||     :::||||                              :::
Db    420 ThrThrThrThrIleThrIleThrThrSerThrThrThrThrThrThr 439

Qy    863 GGCAAGGTGAGATCTTTAATTTAATAGTGCCATCCTTGATAGCTTGTTCAAAATTGAGA 804
             |||        :::||              :::                  |||
Db    440 --------------ThrThrIleThrIleThrThrSerThrThrThrThrThrThrIle 455

Qy    803 ACTTTTCGATTAACTTGGTTCCAATCATCAGCAGCATCCATCTTGGGATCATCATTCA 744
             |||          :::|||              |||                  :::
Db    456 ThrThrThrThrIleThrIleThrThrSerThrThr------------ThrThrThr 472

Db    473 IleThrThrThrThrThrIleThrIleThrThrSerThrThrThrThrThrIleThr 492

Qy    743 ATAGCAGACATGGCTTCAAAAAGACCAAAATAGCTTATCATGAAGTAGTTCCCAACTTT 684
             ||| ||:|||                      |||||                   :::
Db    473 IleThrThrThrThrThrIleThrIleThrThrSerThrThr 492

Qy    683 AATTCTCGACAAGCTTCTTCAAAATCTTGGGTAAATGTCACCCAGTTTGTATTGCTTTTC 624
             |||   :::||                                |||      :::
Db    493 ThrThrIleThrThrThrThrThrIleThrThrThrIleThrThrThr 512

Qy    623 TCCAATTTTTCTGGCATACTGAGCTCCATCCTATThrThrThrIleThrThrThrThrThr 564
             |||                              ||||:||||||||
Db    513 ThrThrThrThrIleThrThrThrThrThrThrIleThrThrThrThrThr 532

Db    552 rThrThrIleThrThrSerThrThrThrThrThrIleThrThrThrThrThrIle 572

Qy    563 ATAACCATTATGCCTACTTCTCCCGTGGTCACCCGGCGCCGCCAGCACTGTTCGGAAGCCAC 504
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 ThrThrThrThrIleThrThrThrThrThrThrIleThrThrThrThrThr 532

Qy    503 G------------------------------------CCGGGAC 503
             |||
Db    533 IleThrIleThrThrThrThrThrThr------ThrThr 552

Qy    502 ----------GAAATAACGAACGGCCGCCT----------CG 483
             |||||||||||                          |||
Db    552 rThrThrIleThrThrSerThrThrThrThrThrIle 572

Db    572 eThrIleThrThrThrThrProLeuProProSerProSerProProSerLeuPr 592

Qy    482 GCCGCGGCGCGCCCCCTCTCCCCCTTCAGCCCAGCGACGGCAGCATGGCTAT 423
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    592 oProProProSerLeuProProProSerLeuProProProLe 612

Qy    422 GCCGGGGAGCGGCGACCCCGCCCCCCCCCCCCCTTACTC 378
             |||||||||||||||||||||||||||
Db    612 uProSerLeuProProProProSerProSerLeuProProProProSerPr 632

Qy    377 TTCCGGCCTCCCCGGCCACCACCCCCTCCCTCG----CCACCGCCTCCGTTCGGCCGATGGCG 321
             |||||||||||                       |||
Db    632 oSerLeuProProProProSerLeuProProProProSerLeuProPr 652

Qy    320 GCCGGCCTTCCTCTCCGCGAGATTAGACCGACGAACCGAAGTTATCTTTTTTTCA 261
             |||||||||||||||||||||||||||||:::
Db    652 oProProSerLeuProProProSerLeuSerProProSerLeu------ 668

Qy    260 CCCTTCGTCGGGAGTCCTGGGTAGTGGGGGGGGGTTCCGGAAGAGAGGGGAGCCCGGAGTCTC 201
             |||||||||||||||||||||||||||||||||||||||||||
Db    669 -ProLeu-------------------------------------- 678

Qy    200 AGAGCCCGGCCCCGTCTGCGCTGCCC 176
             |||                   |||
Db    669 -ProLeu--------------ProProProProSerLeuPr 678

Db    678 oProProProSerLeuProPro 686
```

; PRIOR FILING DATE: 2002-02-19
; PRIOR APPLICATION NUMBER: PCT/CA00/00966
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: US 60/149,466
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: US 60/159,122
; PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 5
; TYPE: PRT
; ORGANISM: Brassica napus
; LENGTH: 721
US-10-086-464-5

Alignment Scores:
Pred. No.:
Score: 142.00
Percent Similarity: 37.04%
Best Local Similarity: 26.85%
Query Match: 2.58%
DB:

Length:	721
Matches:	58
Conservative:	22
Mismatches:	75
Indels:	61
Gaps:	10

US-10-001-857-42 (1-3096) x US-10-086-464-5 (1-721)

```
Qy   788  TGGTTTCAATCATGCCACGATCACTTGGGATCCATCCATTCATTCAATAGCAGACATGGCT  729
            ||| ::  ||||                  ||::
Db     4  TrpSerProCysLeuLeuSerSerProAlaAlaSerPheHisSerLeu------------  19

Qy   728  TCAAAAAGACCAAATAGCTTATCATGAAGTAGTTCTCCCAACTTTTAATTCTCGACAAGCT  669
                                    ||| ::::::     |::::
Db    20  ----------------HisLeuProPheArgSerLysGluAla  29

Qy   668  TCTTCAAAATCTTGGGTAAATGTCACCCAGTTGTAATTGCTTTCTCCCATTTTTTCTCGC  609
            |||                                        :::
Db    30  AspMetSerSer--------------------AlaProSerProGlyThrGly  40

Qy   608  ATACTGAGCTCCATCCTGAATCGTCATCATCTACA-------GAAGTTTCATAACCATT  555
            |||: |||| :|:||                          :::
Db    41  SerProSerProSerAsnSerThrThrThrProProProAlaSerAlaPro  60

Qy   554  ATGCCTACTTCTCCCGTGGTCACCCGACTGTCGGACC----------  506
            |||:||||:||                 ::::  |||
Db    61  ProProThrThrProSer-SerProSerThrIleProThrSerProProPr  80

Qy   505  -ACGGAAATAACGACGCCCGCCCTCCGCGCGCGCCCCTCCCTCAGCCCACCGCCCGA  447
            ||| |                                    :::
Db    80  oSerSerArgSerThrProSerAlaProProProSerProProThrProSerProThrProG1  100

Qy   446  CCGGCAGGTGCACGCATGGTATGGCGGACGGCGACCCCGCC---CCTCTCCCTT  390
            |||                                        :::||||
Db   100  ySerPro-------------ProProLeuProGlnProSerProProAlaProTh  114

Qy   389  ACTCCCGGGGACTTGCGCGTCCCCGGCCACACCCCTCCCACCGGCCTCCGTTC--  332
            ||||:                      |||:|||
Db   114  rThrProGlySerProAlaProValThrProThrArgAsnProProProSerVa  134

Qy   331  -----GGGCGCAATGGGGCGGCCCGCCCTTCTCTTCTCGCGAGATTAGAGACGATCGCGAGACCGGA  276
            |||                                              ::
Db   134  1ProGly-----------ProProSer  139

Qy   275  AGTTATCTTTTTTCACCCTCTGTCGGAGTCCTGGGTAGTGGGCGGGTTCCGGAAAGAGAGC  216
            ||| :||                                        :::
Db   140  ------AsnProSerArgGluGly--------------GlySerProArgProProSe  152

Qy   215  GGAGCCCGGAGTTCGAGAGCCGCGCCCCGGGCTGCGCCCCTGGG  170
            |||:||||                          ::
Db   152  rSerProSerProProSerProSerAspGlyLeuSerThrGly  167
```

; GENERAL INFORMATION:
; APPLICANT: SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH
; APPLICANT: Lloyd, Kenneth O.
; APPLICANT: Yin, Beatrice W.T.
; TITLE OF INVENTION: Nucleic Acid Sequence Encoding Ovarian Antigen, CA125, and Uses
; FILE REFERENCE: 649-A-US
; CURRENT APPLICATION NUMBER: US/10/142,515
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/290,480
; PRIOR FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin version 3.1
;
; SEQ ID NO 11
; TYPE: PRT
; ORGANISM: Human Being
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(5877)
; OTHER INFORMATION: Amino acid sequence of MUC16B
US-10-142-515-11
; LENGTH: 5877

Alignment Scores:
Pred. No.:
Score: 141.50
Percent Similarity: 31.86%
Best Local Similarity: 21.05%
Query Match: 2.57%
DB:

Length:	5877
Matches:	148
Conservative:	76
Mismatches:	266
Indels:	213
Gaps:	27

US-10-001-857-42 (1-3096) x US-10-142-515-11 (1-5877)

```
Qy  2237  CTTCCATTATCCTTCCTCTCGCCATTTGAGAGCCATCGGCACGACTCAATGTTGACATCA  2178
            ||||                                           |||
Db   713  LeuHisLeuAspThrSerThrThrPheMetSerGlyThrHis----------Ser  727

Qy  2177  ACCATGCGTAAAAGAGAATTCAGAGAGATACCAAATATATGTAAATAGTACTCGTGCATACTGT  2118
            |||:                                       ::::
Db   728  ThrAlaSerGlnGlyPheSerHisSerGlnMetThrAlaIleMetSerArgThrProGly  747

Qy  2117  AGAGTTCCAATTCAAAGCCACTTAGAAGGTACTGTATCATCAATAATCCGAAGGTTATGGTAAA  2058
            |||||                           ::::
Db   748  GluValProTrpLeuSerHisProSer-----------ValGluValAlaSerSer  762

Qy  2057  GGACCCAGGTACCTAAACAGGCCAAAATGTTGCCTTTGGGGTTCCTGTTTTCAACAGCATGG  1998
            |||:|||                                    ::::
Db   763  AlaSerPheSerLeuSerSerProValMetThrSerSerSerProValSerSerThrLeu  782

Qy  1997  TGTGAAGCCGCTGCATCAACCTTCTTCGCCTATTAAAATTAAAAAGTGTCATTAAAACTCATCCTG  1938
            ::::
Db   783  ProAspSerIleHisSerSerSerSerLeuProVal------------  793

Qy  1937  CAAGGTGGCAAAATTCCTCAAGAAATATGACCAAGCTTATCTCTCTGTGCAGCCCTGTTATG  1878
            :::||                                        ::::
Db   794  ---------------ThrSerLeuLeuThr----SerGlyLeuValLys  803

Qy  1877  TCCATGGATCTGAATAAGACTACAGAAATGGCCGAAACACAGTGAGTGAACAAAGGAGTCGAT  1818
            :::                                        ::
Db   804  ThrThrGluLeuLeuGlyThrGluThrSerGluProProAsnLeuSer  823

Qy  1817  ACAGTCCTTAGCCGTGATTATTATATTAGGTAGCACTTGGGGAAAGCAACTCGGAGGAT  1758
            |||                                        :::
Db   824  SerThrSerAlaGluIleLeuAlaThrThrGluValThrThrAspThrGlyLysLeuGlu  843

Qy  1757  CTGACAAAAGACCGAAGTGCATCTTTCACCAGTGTCGAGATGAGTTCCAAAGACC  1698
            :::|||                          ||:::
Db   844  MetThrAsnValValThrSerGlyTyrThrHisGluSer----------ProSerSer  859

Qy  1697  TTTTTGTTATCC-------ACCAGGAAAGTGGTTTGTAACAGAGAATCTGAAAGGACACAT  1644
            |||                                        :::
Db   860  ValLeuAlaAspSerValThrThrLysAlaThrSerSerMetGlyIleThrTyrProThr  879

Qy  1643  GGTGACTGTTCACTAAAATTCA----------CAGAAAAATCCAGGATACAATGT  1599
```

RESULT 12
US-10-142-515-11
; Sequence 11, Application US/10142515
; Publication No. US20030078399A1

```
RESULT 13
US-10-243-243A-8
; Sequence 8, Application US/10243243A
; Publication No. US20030104442A1
; GENERAL INFORMATION:
; APPLICANT: Lloyd, Kenneth O.
; APPLICANT: Yin, Beatrice W.T.
; TITLE OF INVENTION: Nucleic Acid Sequence Encoding Ovarian Antigen, CA125, and Use:
; FILE REFERENCE: 649-B
; CURRENT APPLICATION NUMBER: US/10/243,243A
; CURRENT FILING DATE: 2002-09-19
; PRIOR APPLICATION NUMBER: US 10/142,515
; PRIOR FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: PCT/US02/14768
; PRIOR FILING DATE: 2002-05-09
; PRIOR APPLICATION NUMBER: US 60/290,480
; PRIOR FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 5935
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(5935)
; OTHER INFORMATION: Amino acid sequence of MUC16B
US-10-243-243A-8
```

Alignment Scores:

Pred. No.:				
Score:	0.0406		Length:	5935
Percent Similarity:	141.50		Matches:	148
Best Local Similarity:	31.86%		Conservative:	76
Query Match:	21.05%		Mismatches:	266
DB:	2.57%		Indels:	213
	15		Gaps:	27

```
Qy  US-10-001-857-42 (1-3096) x US-10-243-243A-8 (1-5935)

Qy  2237 CTTCCATTATCCTTCCTCTGCCATTTGAGAGCCATCGGACGACTCAAATGTTGACATCA 2178
         |  |||::|||||::|||||
Db   784 LeuThrLeuAspThrSerThrPheMetSerGlyThrHis---------------Ser 798
```

```
Db  1131 IleHisValPheGluAsnThrAlaAlaSerHisValGluThrThrSerAlaThrGluGlu 1150
Qy   566 TTCATAACATTATGCCT------------------------ACTTCTCCGGTTGTCACC 531
         :::|||:::||||||||                        ::|||||::|||:||||
Db  1151 LeuTyrSerSerSerProGlyPheSerGluProThrLysValIleSerProValValThr 1170

Qy   530 CGCGGCGCCACGCACTGTCCGGACCACGGAAATAACGAGCGCGCGCGCGCCC 471
         |||||||||||||||||||||||||||||||||||||||||:|::::: SerSerIleArg
Db  1171 ------------------------------------------SerSerIleArg 1175

Db  1176 AspAsnMetValSerThrThrMetProGly--------SerSerGlyIleThrArgIleGlu 1193
Qy   410 CGACCCGCGCCCCCCCCTCCCTTACTCCCGGGCGCTTCGGCGTCCGGCTCCCCCCTCC 351
         ::|||||                         :::||||
Db  1194 IleGluSerMetSerSerLeuThrProGly------------------------------ 1203

Qy   350 CTCGGCCACGGCCCTCCGGTTCGGCCGATGGCGCGGCCGCCTTCTCTTCGCCGAGGATTAGAGAC 291
         :::|||||||||||||
Db  1204 ---------------LeuArgGluThrArgThrSerGlnAspIleThrSer 1215

Qy   290 GATGCGAGACCGGAAGTTACTTTTTTCACCCTCTGTCGGAGTCCTGGTAGTGGGCGG 231
         |||||||||||||||||||||||||||||||||||||||||||||MetProSerGlyAla
Db  1216 SerThrGluThrSerThrValLeuTyrLys-------------MetProSerGlyAla 1230

Qy   230 TTCCCGGAA 222
         ||||||||
Db  1231 ThrProGlu 1233
```

```
Qy   2177 ACCATGCGTAAAGGAATTCAGAGAGAGATACCAATATATGTAATAGTAACTCGTGCATACTGT 2118
          :::|||::|
Db    799 ThrAlaSerGlnGlyPheSerHisSerGlnMetThrAlaLeuMetSerArgThrProGly 818

Qy        LeuThrArgLysGluSerThrAspMetAlaIle---ThrPro------Lys 1060
Db   1047                                     |||

Qy   2117 AGAGTTCAATTCAAAGCCACTTAGAAGGTTACTGATCATAATGCGGAAGGTTATGGTAA 2058
          |||||||
Db   1076 TCTTCCTCTTCAAAAACACAGCAGCTTTATTTTTTCCCTTTGCAATTGTCACAGANTTTC 1017
          ThrGlyProSerGlyAlaThrSerGlnGlyThrPheThrLeuAspSerThrSerThrAla
Db   1061

Qy   2057 GGACCCAGGTACCTAAACAGGCAAAATGTTGCCTTTGGGTTCCCTGTTTCACAGACATGG 1998
          |||||||                              :::
Db    819 GluValProTrpLeuSerProSer------        ValGluValAlaSerSer 833
Db   1016 AAGATTCCCAGAGCAAAAAGCCTTCATAGCAGGAGGATCTTCTATAAAGTCTGGAATTAATG 957

Qy   1997 TGTGAAGGCGCTGCATCAACCTTCTTCTGCCTTATTAAAATAAATAACTCATCCTG 1938
Db    834 AlaSerPheSerLeuSerSerProValMetThrSerSerProValSerSerThrLeu 853
Db   1081 SerTrpPro-------               GlyThrHisSerAlaThrThrGlnArgPhe 1093

Qy   1937 CAAGGTGGCAAAATTCCTCAAGAATATGACAAAGCTTATCTCTCGTGGAGCCCTGTATG 1878
          |||||||
Db    854 ProAspSerIleHisSerSerSerLeuProVal------   864
Db   1094 ProArgSerValAlaThrThrProMetSerSerArgGlyProGluAspValSerTrpProSer 1113

Qy   1877 TCCATGGATCTGAATAAGACTACACAGAAGTCGGACACAGTGAGTAACAAAAGGATCGAT 1818
          ::|||
Db    865 ------              ThrSerLeuLeuThr---SerGlyLeuValLys 874
Db   1114 ProLeuSerValGlyLeuLysAsnSerProSerSerLeuValSerSerSerValThr 1133

Qy   1817 ACAGTCCTTAGCCTGGTGAATTATTATATAGGTAGCACTGGGGAAAGCACTCGGAGGAT 1758
          |||||||
Db    875 ThrThrGluLeuLeuGlyThrSerSerGluProGlyThrSerSerProProAsnLeuSer 894
Db   1134 SerProSerProLeuTyrSerProSerGlySerSerHisSerSer---     1149

Qy        SerThrSerAlaGluIleLeuLeuAlaThrThrGluValThrThrAspThrGluLysLeuGlu 914
Db   1817

Qy   1757 CTGACAAAAGACCGAAGTGCATCTTTCACCATGTCTTCGGCATGAGATGAGTTCAAAGACC 1698
          :::|||
Db    915 MetThrAsnValValThrSerGlyTyrThrHisGluSer---          ProSerSer 930
Db   1165 ThrAspMetLeuAspAlaSerLeuGluProGluThr---              ThrSerAlaProAsn 1181

Qy   1697 TTTTTTGTTATCC------ACCAGGAAAGTGGTTGTTAACAGAGATCTTGAAAGAACACAT 1644
          |||
Db    931 ValLeuAlaAlaAspSerValThrThrLysGlyIleThrTyrProThr 950
Db   1182 MetAsnIleThrSerArgSerGluSerLeuAlaAlaSerLysAlaThrThrGluThrGlyAla 1201

Qy   1643 GGTGACTGTTCACTAAATTCA------          CAGAAAAATCCAGGATACAAATGT 1599
          |||||||
Db    951 GlyAspThrAsnValLeuThrSerProAlaPheSerAspThrSerArgIleGlnThr 970
Db   1202 IleHisValPheGluAsnThrAlaAlaSerHisValGluThrThrSerAlaThrGlu 1221

Qy   1598 AAATTGTTAAATCACAACCTCACAGACAGTTTTATTCTATTAAATTCTTGCAAAA 1539
          |||||||
Db    971 Lys------                        531
Db   1222 LeuTyrSerSerSerProGlyPheSerGluProThrLysValIleSerProValThr 1241

Qy   1538 TAGTTCACCATTTCTTCCCTTTTAATTATTTTGCAATATCGAGGGAAGGTAGGTGGAAGT 1479
Db   1202 CGCGCGCCGCACGCACTGTCCGGACCACGGAAATAAACGAGCGCGCGCGGCGCGGCGCGGCGCGGCGC 471

Qy   1478 AGCCTCTGGTTCACAAGGGGTTCAAAACCCATCATAATTGGATGATCTCCTTTTGTAGTA 1419
          |||||||
Db    973 LysLeuSerLeuThrProGly------          ThrGlyLysSerThrValLeuSerSerValProThrGlyAla 1006
Db   1247 AspAsnMetValSerThrThrMetProGly---               SerSerGlyIleThrArgIleGlu 1264

Qy   1418 TCATTCTGGGCCTGGATGCCAATGAAGTTAATATGCAAAAGAAGAAGATCTGCT 1359
          |||||||
Db    980 ------                        LeuMetGluThrSerIleSerGluThrSerSer 991
Db   1265 IleGluSerMetSerSerSerLeuThrProGly------     1274

Qy   1358 GCTTGAACATCAATTTGAGCTTCGCAACGCACTGGTCTCTTCTTAGTAAAGGCT 1299
Db    992 Ala------             ThrGlyLysSerThrValLeuSerProThrGlyAla 1006
Db   1275 ------CTCGCCACCGCCTCCGTTCGGCGCGCCGCCTCTCTTCTTCGGAGATTAGAGAC 291

Qy   1298 ATAAGCACTGTCAGTAACACACGAGTAAATTCACTCTGCTGAATACTGCTAAACATTGT 1239
          |||||||
Db   1007 ThrThrGluValSerArgThrGlu------         1014
Db   1287 SerThrGluThrSerThrValLeuTyrLys------          MetProSerGlyAla 1301

Qy   1238 TGGTGTTGTTCAACTTCTGGATCTCTTCTTTCCTTGTCGACTTCGAGTACTCTTT 1179
          |||:|::
Db   1015 ---AlaIleSerSerArgSerIleProGlyProAlaGln--- 1028
Db   1302 ThrProGlu 1304

Qy   1178 ACTCTCTTCTTTGCAATGTTCATCCTCCACATCT------ 1137
          |||||::
Db   1029 -----SerThrMetSerSerAspThrSerMetGluThrIleThrArgIleSerThrPro 1046

Qy   1136 GTAACTCGAAGATCTGTCACACGTTAGCCATTTAAATCCATAAGTCATTGACTGAAAA 1077
Db   1136
```

```
;  APPLICANT:  James E. Bear
;  APPLICANT:  Jurgen Wehland
;  APPLICANT:  Joseph Loureio
;  TITLE OF INVENTION: Methods and Products for Regulating Cell
;                      Motility
;  FILE REFERENCE: M0656/7064 (HCL)
;  CURRENT APPLICATION NUMBER: US/09/823,240
;  CURRENT FILING DATE: 2001-03-30
;  PRIOR APPLICATION NUMBER: 60/194,564
;  PRIOR FILING DATE: 2000-04-03
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: FastSEQ for Windows Version 3.0
;
;  SEQ ID NO 2
;    LENGTH: 802
;    TYPE: PRT
;    ORGANISM: Mus musculus
;  US-09-823-240-2

Alignment Scores:
;  Score:                0.0267
;  Percent Similarity:   140.00
;  Percent Similarity:   36.24%
;  Best Local Similarity: 32.89%
;  Query Match:           9

US-10-001-857-42 (1-3096) x US-09-823-240-2 (1-802)

Length:           802
Matches:           49
Conservative:       5
Mismatches:        58
Indels:            37
Gaps:               6

Qy    535  TCACCCGCGCGCGCCACCACTGTCGGGACCACGGAAATAACGACGCCCGCCCG  476
                 |||| |||  |     |
Db    532  SerProThrProGlnGlyLeuVal----------LeuGlyProProAlaProPro  546

Qy    475  CCGCCCCCTCCCCTCTCAGCCACCGACCGGTGCACGACGTATGCCCGGG  416
                 ||||     ||||||||
Db    547  ProProProProLeuProSerGlyProAlaTyrAlaSerAlaLeuProProProGly  566

Qy    415  ACGCGCGGACCCCGCCCCCTTCCTTACTCCCGGGACTTCGGCGTCCCGGCCCACCC  356
                 |||||      ||| ||
Db    567  ------ProProProProProLeuProSerThrGly-------ProProPro  580

Qy    355  CCTCCCTCGCCACCGCCGCTTCCGTTCGGCGGTCCGCCGCCTTCTTCTTCGCGAGATTA  296
                 ||||
Db    581  ProProProProProProProLeuProAsnGlnAlaProProPro------  595

Qy    295  GAGACGAATCGCGGACCGGAAGTTATCTTTTTCACCCTGTTCGGAGTCCTGGGTAGTG  236
                 |||                    |||  ||  :||
Db    596  ------------ProProProLeuProAlaProSerGlyIlePhe  610

Qy    235  GGCGGTTCCCGGAAAGAG------------GCGGAGCCCGGA  206
                 ||||                       |||
Db    611  SerGlySerThrSerGluAlaAspAsnArgProLeuThrGlyLeuAlaAlaAlaIleAlaGly  630

Qy    205  GTCTCAGAGCCCCGCCCCCCCCTCTCGCCCTCTCGGGAATTTCTTAGAAAGGTGACGGGAA  146
                  :::
Db    631  AlaLysLeuArgGlysValSerArgValGluAspGlySerPhe-------Pro  645

Qy    145  GGCGGAAGGAAAACCGTCGGAGGTTTG  121
                  ::::
Db    646  GlyGly-GlyAsnThrGlySerVal  653

RESULT 15
US-10-196-935A-4
```

```
;  GENERAL INFORMATION:
;  APPLICANT:  Lifton, Richard P
;  APPLICANT:  Wilson, Frederick H
;  APPLICANT:  Choate, Keith
;  APPLICANT:  Ishikawa, Kazuhiko
;  APPLICANT:  Nelson-Williams, Carole
;  TITLE OF INVENTION: COMPOSITIONS METHODS AND KITS RELATING TO TREATING AND DIAGNOSING
;  TITLE OF INVENTION: HYPERTENSION
;  FILE REFERENCE: 04574-5113
```

```
;  CURRENT APPLICATION NUMBER: US/10/196,935A
;  CURRENT FILING DATE: 2002-10-25
;  PRIOR APPLICATION NUMBER: US 60/306,084
;  PRIOR FILING DATE: 2001-07-17
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: Patentin version 3.1
;
;  SEQ ID NO 4
;    LENGTH: 1243
;    TYPE: PRT
;    ORGANISM: Homo sapiens
;  US-10-196-935A-4

Alignment Scores:
;  Pred. No. :
;  Score:                  0.0341
;  Percent Similarity:     139.50
;  Percent Similarity:     35.24%
;  Best Local Similarity:  29.05%
;  Query Match:            2.54%

US-10-001-857-42 (1-3096) x US-10-196-935A-4 (1-1243)

Length:           1243
Matches:            61
Conservative:       13
Mismatches:         63
Indels:             73
Gaps:               10

Qy    535  TCACCCGCGCGCGCCACCACTGTC-----CGGACCACGGAAATAACGACGCCCGCCCTCGG  482
                 |||  |||                    :||                  |||
Db      4  SerProAlaThrGlyThrThrValLeuMetSerGlnThrGlyHisAlaAspLeuLeuArg  23

Qy    481  CCGCCCCGCCGCCCC------------CCTCCCCTCTCAGCCCACCCAGGACCGGCAG  440
                 |||||||||||                |||                  |||
Db     24  ProProProProLeuProGlyThrAlaGlyGlnProLeuGlyProProProAlaProProAlaArgAla  43

Qy    439  CGT------------GCACGCATGCGGTAATGCCCGGGACGCGC--------  410
                 |||
Db     44  ArgArgPheSerGlyLysAlaGlyProArgProArgSerSerArgLeuSerArgSer  63

Qy    409  ----------------------GACCCCGCCCCCCCCTCTCCC  392
                 
Db     64  SerValAspLeuGlyLeuLeuSerSerTrpSerLeuProAlaSerProAlaProAspPro  83

Qy    391  ----------------TTACTCCCGGGACTTCGGCGTCCCGGCCCCACCACCCCCTCCC  350
                 :||   :||
Db     84  ProAspProProAspSerAlaGlyProGlyProAlaArgSerPro-----ProProSer  101

Qy    349  TCGCCACCGCCGCTTCCGTTCGGCGGATGGGGCGGCCTTCTTCTTCGCGAGATTAGAGACG  290
                 :::
Db    102  SerLysGluProGroGluGlyThrThrTrpThrGluGlyAlaProValLysAlaAlaGluAsp  121

Qy    289  ATCGCGAGACCGGAAGTTATCTTTTTCACCCTGTTCGGAGTCCTGGGTAGTGGCGGT  230
                 ||||||||  |||
Db    122  SerAlaArgProGlu----------LeuProAspSerAlaValGlyProGly  135

Qy    229  TCCCGGAAAGAGGCGGGAGCCCGGAGAGTCTCAGAGCCCCGTCTGCCCTCTCTGGG  170
                 |||
Db    136  SerArgGlu---------------------------ProLeuArgValProGluAla  145

Qy    169  AATTTCTTAGAAAGGTGACGCGAAGCGGAAGAAAACCGTGGAGGTTTGCGTAGACCTTA  110
                 |||||||   |||
Db    146  ValAlaLeuGluArgGlyArgArgGlu-------------------------------  153

Qy    109  GTGGTTGGGAGCAGCAGAACGAGCAGCCAAGCTCCGAGGAGCAGCAACAGCGTGCTACTGCTGCCT  50
                 ||:||              |||:||
Db    154  ------GlnGlnGlyLysGluAspMetGluThrThrSerPro  169

Qy     49  GACAACGGATTRATG---TTTCCTATACAG  23
                 ||:||   :::||
Db    170  AspGlyArgTyrLeuLysPheAspIleGlu  179
```

```
OM nucleic - nucleic search, using sw model

Run on:             November 24, 2003, 21:46:55 ; Search time 6196 seconds
                                                  (without alignments)
                    GenCore version 5.1.6           12144.393 Million cell updates/sec
                    Copyright (c) 1993 - 2003 Compugen Ltd.

Title:              US-10-001-857-42
Perfect score:      3096
Sequence:           1 ttcctcacgaaactccagg..............atttcttgacaaaaaaaa 3096

Scoring table:      IDENTITY NUC
                    Gapop 10-0 , Gapext 1.0

Searched:           22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters:        45562784

Post-processing:    Minimum Match 0%
Maximum DB seq length: 2000000000
Minimum DB seq length: 0
Maximum Match 100%
Listing first 45 summaries

Database :          EST:*
          1:  em_estba:*
          2:  em_esthum:*
          3:  em_estin:*
          4:  em_estmu:*
          5:  em_estov:*
          6:  em_estpl:*
          7:  em_estro:*
          8:  em_htc:*
          9:  gb_est1:*
         10:  gb_est2:*
         11:  gb_htc:*
         12:  gb_est3:*
         13:  gb_est4:*
         14:  gb_est5:*
         15:  em_estfun:*
         16:  em_estom:*
         17:  em_estin:*
         18:  em_gss_hum:*
         19:  em_gss_inv:*
         20:  em_gss_pln:*
         21:  em_gss_vrt:*
         22:  em_gss_fun:*
         23:  em_gss_mam:*
         24:  em_gss_mus:*
         25:  em_gss_pro:*
         26:  em_gss_rod:*
         27:  em_gss_phg:*
         28:  em_gss_vrl:*
         29:  gb_gss1:*
         28:  gb_gssI:*
         29:  gb_gss2:*
```

                                              SUMMARIES

```
                                  %        Query
Result No.   Score   Match   Length   DB ID        Description
---------   -----   -----   ------   -----        -----------
    1        795    25.7    1033     11  BC027201   BC027201 Mus muscu
    2        777.8  25.1    973      11  AK021042   AK021042 Mus muscu
c   3        705    22.8    722      12  BM979022   BM979022 UI-CF-DU1
    4        696.8  22.5    854      10  BF969365   BF969365 60227l630
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
c    5  658.6   21.3    814    14  CB169246   CB169246 VBB603020
c    6  653.8   21.1    3071   11  AK050805   AK050805 Mus muscu
     7  653.6   21.1    768    13  AK050805   BG547889 62575187
     8  648    20.9    957    13  BQ965085   BI602827 60347187
c    9  641.4   20.7    828    9   AI790514   BQ965085 AGENCOURT
c   10  610     19.7    616    12  BM979512   AI790514 ul02b03.x
c   11  599.2   19.4    900    14  CA977646   BM979512 UI-CF-DU1
    12  598.8   19.3    697    12  BI602827   CA977646 AGENCOURT
    13  597     19.3    598    9   AW007583   BI602827 60347187
    14  593.8   19.3    718    13  BU058931   AW007583 wt02f11.x
c   15  586     18.9    972    10  BG112812   BU058931 UI-M-FR0-
    16  584.8   18.9    2958   11  BC043701   BG112812 602285005
    17  575.2   18.6    763    13  BU613022   BC043701 Mus muscu
c   18  572.4   18.5    952    13  BQ925670   BU613022 UI-M-FR0-
c   19  560     18.1    560    9   AA888870   BQ925670 AGENCOURT
    20  555.6   17.9    927    11  BC010475   AA888870 ak23f02.s
    21  551.2   17.8    1117   13  BU525571   BC010475 Mus muscu
    22  548.4   17.7    873    9   AU081000   BU525571 AGENCOURT
    23  542     17.5    542    9   AU081000   AU081000 AU081000
    24  538.6   17.4    735    13  BU057675   BU057675 UI-M-FR0-
    25  537     17.3    537    12  BF468052   BF468052 UI-M-CD0-
    26  526.8   17.0    666    13  BM69943    BM69943  UI-E-DX1-
    27  523.2   16.9    637    14  BQ43173    BQ43173  UI-M-EV0-
    28  507.8   16.4    509    10  CD372119   CD372119 UI-R-G00-
    29  506.8   16.4    906    12  BI658829   BG612374 60264O115
    30  503.2   16.3    1114   10  BG245408   BI658829 603301980
    31  499.6   16.1    640    13  BQ443252   BG245408 602254951
    32  496     16.0    1049   11  BQ885743   BQ443252 UI-M-EV0-
    33  495.6   16.0    631    14  CB418817   BQ885743 AGENCOURT
    34  491.8   15.9    690    10  AW975643   CB418817 591640 MA
    35  491.4   15.9    901    12  BI731865   AW975643 EST387752
    36  490.8   15.7    494    12  BM510376   BI731865 603353606
    37  487     15.6    721    9   AJ453000   BM510376 i140g03.y
    38  482.2   15.6    615    14  CA749517   AJ453000 AJ453000
    39  480.2   15.5    479    9   AI339059   CA749517 UI-M-FY0-
    40  472.6   15.3    485    9   AW090031   AI339059 qq29e09.x
    41  470.4   15.2    480    14  CB242727   AW090031 xd01a08.x
    42  465     15.0    556    10  BG062635   CB242727 UI-CF-FN0
    43  464.8   15.0    558    14  CA555718   BG062635 L095SF03-
    44  463.2   15.0    569    12  BI682349   CA555718 K0201G09-
    45  461.6   14.9    569    12  BU260643   BI682349 463408 MA
        461.4   14.9    784    13  BU260643   BU260643 603503268
```

                                              ALIGNMENTS

Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: amg@bcm.tmc.edu
Gunaratne, P.H.; Garcia, A.M.; Lu, X.; Hulyk, S.W.; Hale, S.M.,
Gunaratne, P.H.; Garcia, A.M.; Lu, X.; Hulyk, S.W.; Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S.; Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 31 Row: i Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: Similarity but not
identity to protein
This clone has the following problem: frame shifted.

FEATURES   Location/Qualifiers
 source   1..1033
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="FVB/N"
    /db_xref="taxon:10090"
    /clone="IMAGE:4483189"
    /tissue_type="Mammary tumor. Metallothionien-TGF alpha
    model. 10 month old virgin mouse. Taken by biopsy."
    /clone_lib="NCI_CGAP_Mam1"
    /lab_host="DH10B"
    /note="Vector: pCMV-SPORT6"

BASE COUNT 303 a 216 c 272 g 242 t
ORIGIN

Query Match   25.7%; Score 795; DB 11; Length 1033;
Best Local Similarity 88.3%; Pred. No. 4.2e-133;
Matches 864; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Qy 541 GGAGAGTAGGCATAAATGGTTATGAAAGCTTCTGTAGATGATGACATTCAGGATGGGAG 600
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 54 GGCGCGGACGCGGCATCATGGTTATGAAAGCTTACTGTAGACGACGATGCTTCGGGATGGGAG 113

Qy 601 CTCAGTATGCCCAGAAAAATGGAGAAAAAGCAATACAAAACTCGGGTGGGACATTACCCAAGAT 660
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 114 CTCGGGGTCCCCGGAAAAAAATGGAAAAAGTAGCACAAAGCTGGGTGGACATAACCCAGGAT 173

Qy 661 TTTGAAGAAGCTTGTCGTGAGGAGAATTAAAAGTTGGGAGAACTACTTCATGATAAGCTATATTGGT 720
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 174 TTTGAAGAGTCTTGTGTCGAGAGCTCAGAGTTGGGAGAACTGCTTCATGATAAGCTGTTTGGT 233

Qy 721 CTTTTTGAAGCCATGTCTCGCTATTGAAAATGATGGAATCCCAAGATGATGGCTCGGCATCATT 780
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 234 CTTTTTCGAAGCCAATGTCTCGCTATTGAAAATGATGGAATCCTAGGATGGGATGCCGGTATGATC 293

Qy 781 GGAAACCAAGTTAATCGAAAAGTTCTCAATTTTGAACAAGCTATCAAGGAATGGCACTATT 840
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 294 GGGAACCAAGTAGAAACAGAAAAGTTCTCAAGTTTGAACAAGCTATCAAGGATGGCACCATT 353

Qy 841 AAAAATTAAAGATCTCACCTTGCCTGATTAGGGATTATCGATACAACATTGTTTTTGCTGT 900
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 354 AAAAATTAAAGACCTCAGCCTGCCTGAACTGAATAGTTATGGACACCGTTTCTGCTGT 413

Qy 901 TTGATAACGTGTGGTTAGAACGGCCATTCACTGGCACAGACAGTATTTACGTGCCTTTACATT 960
     |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 414 TTGATCACATGGCTCGAAGGCCATTCCTTGGCACAGACAGTGTTTACGTGCCTTTACATT 473

Qy 961 CATAAATCCAGACTTTATAGAAGAATCCTGCTATGAAGGCTTTTGCTCTTTTTGAAAA 1020
     |||||||||||||||||||||||||||||||||||||||||||||||||
Db 474 CATAAATCCCGACTTTCATAGAAGATCCTGCCATGAAAGCTTTTGCTCTTGGGGAATCTTGAAG 533

Qy 1021 ATCTGTGACATTGCAAGGGAAAAAGTAAATAAATAAAGCTGCGTGTTTTGAAGAGGAAGAGTTT 1080
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 534 ATCTGCCGACATTGGAGGAAAAAGTAAATAAAAGTAAAGGCAACAGTGTGACAAGATCTCCGAGTTACAGG 593

Qy 1081 CAGTCAATGACACTTATGGATTTTAAATTGGGTTAACAGTGTGCAGATCTTCGAGTTACAGGC 1140
     |||||||||||||||||||||||||||||||||||||||||||||
Db 594 CAGTCAATGACACTACGGGATTTAAATGGCCAACAGTGTGACAGATCTCCGAGTTACAGG 653

Qy 1141 ATGCTAAAAGATGTGGGAGGGATGACAATGCAAAGAAGAGAGAGTAAGAGTACTCGAGTCGACAA 1200
     |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 654 ATGCTAAAGGATGTGGGAAGATGATCTGCAAAAGAAGAGAGTAAGAGTACTCGAAGTCGACAA 713

Qy 1201 GGAGAAGAGAAGAGGATCCAGAAGTTGAACTAGAACACACCAACAATGTTTAGCAGTATTCAGC 1260
     ||||||||||||||||||||||||||||||||||||||||||||||
Db 714 GGAGAAGAGAAGAGAATCCAGAAGTTCGAACTAGCCAACTAGAAC 773

Qy 1261 AGAGTGAAAATTACTCGTGTTACTGACAGTGCTTATAGCCTTTACTTACTTAAGAAAGAAGACC 1320
     |||||||||||||||||||||||||||||||||||||||||||||||||
Db 774 AGAGTGAAGTTCACCGCGAGTGCTTCAACAGTGCTTCATAGCCTTTACTTAAGAAAGAAGACC 833

Qy 1321 AGTGCTGTTGCAGAGATCCAAAAATTGATGGTTCAAGCAGCAGAGATCTTCTTTCTGCCATT 1380
     ||||||||||||||||||||||||||||||||||||||||||||||||
Db 834 AGTGCTGTTGCAGAGGCTCAGAAACTGATGGTTCCAGGCAGGAAACTTCCAGCAGTTTTCTGCCATT 893

Qy 1441 ATTATGATGGGTTTTGAACCCTTGTGAGAGGTACTTCCACTTCCCTGA 1500
     |||||||||||||||||||||||||||||||||||||
Db 894 CACACCTCATTGCACCCACGAGCTCAGGCTACCAAAGGAGGACCATCCA 953

Qy 1441 ATTATGATGGGTTTTGAACCCTTGTGAGAGGCTACTTCCACTTCCCTTACCTTCCCTGA 1500
     ||||||||||||||||||||||||||||||||||||
Db 954 ATTATGATGGGTTTTGAGCCCTTGTTAACCAGAGATTACTTCCACCCACCTTCCCTCGC 1013

Qy 1501 TATGCAAAAATAATTAAAA 1519
     ||||||||||||
Db 1014 TATGCAAAAATAATTAAAAAAAAAAAAA 1032

RESULT 2
AK021042
LOCUS   AK021042 973 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male corpus striatum cDNA, RIKEN full-length
   enriched library, clone:C030004C14 product:CORNEAL WOUND HEALING
   RELATED PROTEIN homolog [Rattus norvegicus], full insert sequence.
ACCESSION AK021042
VERSION AK021042.1 GI:12861810
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
   Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
   prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
   Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
   Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
   Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
   Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
   Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
   Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
   sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
   Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
   Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,

AUTHORS   Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C.A., Wilming,L.,
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.,
and Hayashizaki,Y.

TITLE   Functional annotation of a full-length mouse cDNA collection

JOURNAL   Nature 409 (6821), 685-690 (2001)

MEDLINE   21085660

PUBMED   11217851

REFERENCE   5 (bases 1 to 973)

AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.

TITLE   Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

JOURNAL   Nature 420, 563-573 (2002)

REFERENCE   6 (bases 1 to 973)

AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
Tejima,Y., Toya,T., Yamamura,Y., Yasunishi,A., Yoshida,K.,
Yoshino,M., Muramatsu,M. and Hayashizaki,Y.

TITLE   Direct Submission

JOURNAL   Submitted (18-AUG-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

COMMENT   Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGAGAGAGAGACTCGTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through two rounds of normalization to Rot = 20.0 and subtraction
to Rot = 478.8. Second strand cDNA was prepared with the primer
adapter of sequence [5'
GAGAGAGAGATTCTGGAGTTAATTAATCCCCCCCCCCC 3']. cDNA was cleaved
with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3'
end: BamHI. Host: DH10B.

FEATURES   Location/Qualifiers
     source   1..973
          /organism="Mus musculus"
          /mol_type="mRNA"
          /strain="C57BL/6J"
          /db_xref="FANTOM_DB:C030004C14"
          /db_xref="MGI:1912689"
          /db_xref="taxon:10090"
          /clone="C030004C14"
     misc_feature   24..>971
          /note="CORNEAL WOUND HEALING RELATED PROTEIN homolog
          [Rattus norvegicus] (SPTR|Q9JI01, evidence: FASTY,
          96.2%ID, 43.5%length, match=948)
          putative"
          /db_xref="MGI:1925939"
          /sex="male"
          /tissue_type="corpus striatum"
          /clone_lib="RIKEN full-length enriched mouse cDNA library"
          /dev_stage="adult"

BASE COUNT   285 a   202 c   250 g   236 t
ORIGIN

Query Match   25.1%;   Score 777.8;   DB 11;   Length 973;
Best Local Similarity   87.9%;
Matches 848;   Conservative 0;   Mismatches 117;   Indels 0;   Gaps 0;

```
Qy  541 GGAGAAGTAGCCATAATGGTTATGAAAAAGCTTCTGTGAGATGAATGACGGATTCAGGATGGGGAG 600
            |||||||||| |||| ||||||||||||  ||||| ||||| ||| ||| |||||||||| ||
Db    9 GGCGGGCCTAGCGACTCATGGTTATGAAAAAATGGGTTATGAAAGACCGAGCGATGCTTCGGATGGGAG 68

Qy  601 CTCAGTATGCCAGAAAATAAATGGAGAAAAGCAATACAAACTGGGGTGGACATTACCCAGAT 660
            ||||  ||||||| ||||||||||||||||||||
Db   69 CTCGGGGGCTAGCGACTCATGGTTATGAAAAAATGGAAGTTACTGTGAGCAGACGATGCTTCGGATGGGAG 128

Qy  661 TTTGAAGAAGCTTGTCCAGAATTAAAGTTGGGGAGGAGAACTTCATGATAAGCTATTTGGT 720
            ||||||| ||||||||||||||||||||||||||||
Db  129 TTTGAAGATGCTTGTCGTGCAGAGCTGAAAGTTGGGGACGAACTGCTTCATGATAAGCTGTTTGGT 188

Qy  721 CTTTTTGAAGCCATGTCTGCTCTATTGAAATGATGGAATGGATGGGGCCATGATT 780
            ||||||||| | |||||||||||| |||||||||||||||||||||||||||
Db  189 CTTTTTGAAGCCATGTCTGCTCTATTGAAATGATGGAATGGATGGGGCCGGTATGATC 248

Qy  781 GGAAACCAAGTTAATGAAAAGTTCTCAAATTTTGAAACAAGCTATCAAAGGATGGGCACTATT 840
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db  249 GGGAACCAAGTGAAATAAGAAAAGTTCTCAAATTTTGAACAAGCTATCAAAGGATGGCCACATT 308

Qy  841 AAAAATTAAAGATCTCACTTGCCTGAAACTGATCAGGAATTATGGATACATGTTTTTTGGTGT 900
            |||||||||||||||| ||||||||||||||| |||||||||||||||| |||||
Db  309 AAAAATTAAAAACCAGCCTGCCTGACCTGATAGGAATAATGAACACCTGTTTCTGTGT 368

Qy  901 TTGATAACGTGGTTAGAAGGCCATTCACTGGCACAGACAGTATTTACGTGCCTTTACATT 960
            ||||||||||||||||||||||||| |||||||||||||||||||||||| ||||
Db  369 TTGATCACATGGCTCGAAGGCCATTCCTTGGCACAGACAGGGTTTACGTGCCTTTACATT 428

Qy  961 CATAATCCAGACTTTATAGAAGATCCTGCTATGAAAGGGTTTTGCTCTGGGAAATCTTGAAA 1020
            ||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db  429 CATAATCCCGACTTCATAGAAGATCCTGCTATGAAAGGCTTTTGTTCTGGGAATCTTGAAG 488

Qy 1021 ATCTGTGACATTGCAAGGGAAAAAGTAAATAAAGCTGCGTGTTTTGAAGAGGAAGATTT 1080
            |||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  489 ATTTGGTACATTGCACGGGAAAAAGTAAATAAAGCTGCGTGTTTTCGAAGAGGAAGATTTC 548

Qy 1081 CAGTCAATGACTTATGGATTTAAAATGCTAACAGTGTCAACAGACGTTACAGGC 1140
            ||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  549 CAGTCAATGACTACGGATTTAAAATGCCAACAGTGTCAACAGATCTCCGAGTTACAGGG 608

Qy 1141 ATGCTAAAAGATGTGGAGGAGGATGAACATGCCAAAGAGAGGAGGTAGCGAAGTCGACAA 1200
            |||||||||||||||| ||||||||||||| ||||||||||||||| |||||||||||
Db  609 ATGATAAAGGATGTGAGAGATCATCCGAAAGGCGAGTAGGAGTACTTGAAGTCGACAA 668

Qy 1201 GGAGAAGAGAAGAGACCAGAAGTTGAACTAGAACACCAACAATGTTAGCAGTATTCAGC 1260
            |||||||||||||||||||||||||||| |||||||| |||||||||| ||||||
Db  669 GGAGAAGAGAGATCCAGAGGTCGAACTAGAACACCAGCAGGCGTTGGCAGCATTCAGC 728

Qy 1261 AGAGTGAAATTTACTCGTGTGTTTACTGACAGTGCTTATAGCCTTTACTAAGAAAAGGACC 1320
            |||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db  729 AGAGTGAAGTTCACCGGAGTGCTGCTCACCAGTGCTTGCTCAGCGCCTTTACTAAGAAAAGAGACC 788

Qy 1321 AGTGCTGTTGCAGAGGCCTCAAAAATTGATGGTTCAAGCAGAGCAGATCTTCTTTCTGCCATT 1380
            ||||||| |||||||||||||||||||||||||||||||| |||||||||||||||||
Db  789 AGTGCTGTTGCAGAGGCTCAGAAACTCGATGGTTCCAGGCGCAGGGCGCAGACTTTCTTTCGCCATT 848
```

```
Qy   1381 CATAATTCATCATCATGGCATCCAGGCCCAGATAGTGATAGTACTAACAAAAGGAGAGATCATCCA 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    849 CACACCTCATTGCACCACGGCATCCAGGCTCAGAATGGCACTACCAGGAGGACCATCCA 908

Qy   1441 ATTATGAGTGGGGTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACCTTCCCTCGA 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    909 ATTATGAGTGGGGTTTGAGCCCTTGTTAACCAGAGATTACTTCTCCACCACCTTCCCTCGC 968

Qy   1501 TATGC 1505
          |||||
Db    969 TATGC 973
```

RESULT 3
BM979022
LOCUS   BM979022/c   722 bp   mRNA   linear EST 21-FEB-2003
DEFINITION UI-CF-DUI-adl-c-13-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone
            UI-CF-DUI-adl-c-13-0-UI 3', mRNA sequence.
ACCESSION   BM979022
VERSION   BM979022.1 GI:19599047
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 722)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE   Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   MEDLINE 97044477
   PUBMED   8889548
COMMENT   Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com) or from Open Biosystems
            (www.openbiosystems.com).
            Seq primer: M13 FORWARD
            PolyA=Yes.
FEATURES   Location/Qualifiers
     source   1..722
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-DUI-adl-c-13-0-UI"
                     /tissue_type="Primary Lung Epithelial Cells"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-CF-DUI"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site 1: EcoR I; Site 2: Not I;
                     UI-CF-DUI is a normalized cDNA library containing the
                     following tissue(s): Primary Lung Epithelial Cells The
                     library was constructed according to Bonaldo, Lennon and
                     Soares, Genome Research, 6:791-806, 1996. First strand
                     cDNA synthesis was primed with an oligo-dT primer
                     containing a Not I site. Double stranded cDNA was ligated
                     to an EcoR I adaptor, digested with Not I, and cloned
                     directionally into pT7T3-Pac vector. The oligonucleotide
                     used to prime the synthesis of first-strand cDNA contains
                     a library tag sequence that is located between the Not I
                     site and the (dT)18 tail. The sequence tag for this
                     library is GGCTGTAGGC.
                     TAG_LIB=UI-CF-DUI
```

BASE COUNT     215 a    147 c    124 g    236 t
ORIGIN
```
Qy   Query Match   Best Local Similarity   22.8%;   Score 705; DB 12; Length 722;
Db   722 Matches 716; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
                                          99.9%;   Pred. No. 7.2e-117;

Qy   2380 CAAAGTAGGTAAAAGGCGAAGTTTGATGTTTGACTTGATAGTGAAGTTCGGTATGAAGAACACAGGTT 2439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   722 CAAAGTAGGTAAAACCGGAAGTTTGAGTTGAGTGTGAAGAACAGTTCGGTATGAAACACAGG-T 664

Qy   2440 TGCTCCAATCCAACACAGTGTGATGAGGACCCGGCCAGTGCCACTACTTACAGTTCAAGGAAAT 2499
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   663 TGCTCCAATCCAACACGTGTGATGAGGACCCGGCCACTACTTACAGTTCAAGGAAAT 604

Qy   2500 GTCTGACCTCAATAAATATAGCCCTCCTCGTCTCTCGAACTGTATGTGCAGCTAG 2559
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   603 GTCTGACCTCAATAAATATAGCCCTCCTCGTCTCTCGAACTGTATGTGCAGCTAG 544

Qy   2560 TAAGCACTTTCAACAGGCAAAAAAATGAATATGATAATATTCCTAAACCCGGACCATGAGGT 2619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   543 TAAGCACTTTCAACAGGCAAAAAAATGATATTGGAAAATATTCCTAAACCCGGACCATGAGGT 484

Qy   2620 TAATAGAATTTTAAAGGTTGGCAAACCAAACTTGTGTGGTTATGAAGTTATTGGCAGGAGG 2679
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   483 TAATAGAATTTTAAAGGTTGGCCAAAACCAAACTTGTGTGGTTATGAAGTTATTGGCAGGAGG 424

Qy   2680 ACACAAAAAAGGAATCTAAAGTTCCTCCTGCTCATAAATATATTTCC 2739
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   423 ACACAAAAAAGGAATCTAAAGTTCCTCCTGCTCATAAATATTTCC 364

Qy   2740 TGTTGTGAAACTTGTTTGAGAGAGACTGGGCCGAGGTGGGCCATAAAGGGGCAGAGTCTTCTT 2799
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   363 TGTTGTGAAACTTGTTTGTTGAGAGAGACTGGGCCGAGGTGGGCCATAAAGGGGCAGAGTCTTCTT 304

Qy   2800 TCAGACCCAACTCTTAGAGGGGCCAGCTCCACTCACGGGAGAGTGAGATGGAT 2859
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   303 TCAGACCCAACTCTTAGAGGGGCCACATCACCACGGGAGAGTGAGATGGAT 244

Qy   2860 TTCTTGGGTAACAACTCATTATAAGGAATACTTTAGTTTGACAGCCTTATATGACATGA 2919
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 TTCTTGGGTAACAACTCATTATAAGGAATACTTTTAGTTTGACAGCCTTATATGACATGA 184

Qy   2920 ATGAAAACTGCTGTTTTAAAGTGGTTTTATATATGTTCCATGGAAGAGAAAACTGGTCTTATTGA 2979
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   183 ATGAAAACTGCTGTTTTAAAGTGGTTTATATATGTTCCATGGAGAGAAAACTGGTCTTATTGA 124

Qy   2980 ATGCATTGATGAACGTTATATGGTTTTTTATTACAGATTTAATCACAAATCATTTTTATGA 3039
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123 ATGCATTGATGAACGTTATATGGTTTTTTATTACAGATTTAATCACAAATCATTTTTATGA 64

Qy   3040 ATGATTGAGTGAAAAATAGTGTTTTATAAAGTTAATAAAATTTCTTGACAAAAAAA 3096
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 ATGATTGAGTGAAAATAGTGTTTTATAAAGTTAATAAAATTTCTTGACAAAAAAA 7
```

RESULT 4
BF969365
LOCUS       BF969365               854 bp    mRNA    linear  EST 22-JAN-2001
DEFINITION  602271630F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4357702 5',
            mRNA sequence.
ACCESSION   BF969365
VERSION     BF969365.1  GI:12336580
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 854)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)

TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
TAG_SEQ=GGCTGTAGGC"
TAG_LIB=UI-CF-DUI

```
JOURNAL    Unpublished
COMMENT    Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM9999 row: 1 column: 07
           High quality sequence stop: 785.
FEATURES            Location/Qualifiers
     source         1..854
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:4359702"
                    /tissue_type="adrenal cortex carcinoma, cell line"
                    /lab_host="DH10B (phage-resistant)"
                    /clone_lib="NIH_MGC_84"
                    /note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
                    NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT
                    primed. Average insert size 1.229 kb. Library enriched for
                    full-length clones and constructed by Life Technologies.
                    Note: this is a NIH_MGC Library."
BASE COUNT      273 a    158 c    186 g    237 t
ORIGIN

Query Match        22.5%;  Score 696.8;  DB 10;  Length 854;
Best Local Similarity   94.5%;  Pred. No. 2.1e-115;
Matches 809;  Conservative    0;  Mismatches  37;  Indels  10;  Gaps   8;

Qy       2245 GAAAGGCCGTAGTAGTAAAAAAAAACAAAG-AAAAAAAAGAAAAGTTCGCCCATTGAGCCGAG 2303
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db          1 GAAAGGCCGTAGTAGTAAAAAAAAACAAAAAAAAAAAAGAAAAGTTCGCCCATTGAGCCGAG 60

Qy       2304 AGATCACAATGAGCCAAGCACATATCAGACACATCAGACATGTGTGCTGGGAATGGTAG 2363
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGATACAATAGGCCAAGCATATCAGACAACATGTGTGCTGGGAATGTTAAAACCATGGTAG 120

Qy       2364 CATTTGACATGGACGCGCAAAGTACGTAAACCGGAAGTTTTGACTTGAGTGAACAAGTTC 2423
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CATTTGACATGGACGGGCAAAGTACGTAAACCGGAAGTTTTGAGCTTGATTAGTGAACAAGTTC 180

Qy       2424 GGTATGAACACAGGTTTGCTCCATTCAACAGTGTGATGACCCGGTCGCACTACT 2483
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGTATGAACACAGGTTTGCTCCATTCAACAGTGTGATGACCCGCCCAGTGCACTACT 240

Qy       2484 TACAGTTCAAGGAAATCTCTGACCTCAATAAATATAGCCCCTCCTCCAGTCTCCTGAAC 2543
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TACAGTTCAAGGAAATGTCGACCTCAATAAATATAGCCCCTCCTCCCAGTCTCCTGAAC 300

Qy       2544 TGTATGTGTGGCAGCTAGTTAAGACATTTCAACAGCAGAAAAATGATATTGGAAAATATTCCTA 2603
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGTATGTGGCAGCTAGTTAAGCACTTTCAACAGCCAATAAATGATATTGGAAAATATTCCTA 360

Qy       2604 ACCCGGACCATGAGGTTAATAGAATTTAAGGTTGCCAAAACCCAACTTTGGTGGTTATGA 2663
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ACCCGGACCATGAGGTTAATAGAATTTTAAGGTTGCCAAAACCCAACTTTGTGGTTATGA 420

Qy       2664 AGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCTCTGAATTTGATTTCTCTG 2723
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGTTATTGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCTCTGAATTTGATTTTCTCTG 480

Qy       2724 CTCCAAAAATATTTTTCCTGGTTGTGAAACTTGTTTGAGGAGGACTGGGGAGTGGCCAATAAA 2783
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTCATAAAATATTTTTCCGTGTTGTGAAACTTGTTTGAGAGAGACTGGGGAGTGGCCAATAA 540

Qy       2784 GGGGCAGAGAGTCTTCTTTCAGACCCAACTCTTAGACACCACCAGGCTCCACATCAC 2843
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGGGCAGAGAGTCTTCTTTCAGACCCAACTCACCAGGCTCCACATCAC 600
```

```
Qy       2844 GGGAAGTGAGATGAGATTTCTTGGGTAACAACTCATTATAATAAGGAATACTTAG-TTTTGAC 2902
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        601 GGGAAGTGAGATGAGATTTCTTGGGTAACAACTCATTATAATAAGGAATACTTTAGTTTTGAC 660

Qy       2903 AGCCTTATATGACA-TGAATGAAAAACTGCTGTTTAA-AGTGGTTATTATGTTCCAT-G 2959
              |||||||||||||| ||||||||||||||||||||| ||||||||||||||||||| |
Db        661 AGCCTTATATGACATTGAATGAAAAACTGCTGTTTAACAGTGGTTATTATGTTCCCTGG 720

Qy       2960 GAAGAAACTGGTCTTTATTGAATGCATTGATGAAGTTTATTATGTTTATTACAGATTTAA 3019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GACGAAACTGGTCTTTATTGAATGGCTTGATGAACG-TATATGGGTGTTATTACCGATTTA- 778

Qy       3020 TCACAAATCATTTTTTATGAATGATTGAGTGAAAAATAGTTGTTTATAAAGGTTAATAAAAT 3079
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 --TCCCAATCATTTTATGAATGATTG-GTGCACATCGGTTTAACGTGATAAATTCTTG 835

Qy       3080 TCTTGACAAAAAAAAA 3095
              || |||||||
Db        836 CGCAAAAAACAAAAAA 851
```

```
RESULT 5
CB169246/c
LOCUS       CB169246                814 bp    mRNA    linear   EST 30-JAN-2003
DEFINITION  VBB603020215.R1 CSEQFXN41 testes Bos taurus cDNA, mRNA sequence.
ACCESSION   CB169246
VERSION     CB169246.1  GI:28155372
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bovini; Bos.
REFERENCE   1  (bases 1 to 814)
  AUTHORS   Adelson,D.L. and Gill,C.A.
  TITLE     Bovine ESTs (Adelson and Gill)
  JOURNAL   Unpublished
COMMENT     Contact: David L. Adelson
            Animal Breeding and Genetics
            Texas A&M University
            Animal Science Dept., TAMU-2471, College Station, TX 77843-2471,
            USA
            Tel: 9798452616
            Fax: 9798456970
            Email: david.adelson@tamu.edu.
FEATURES            Location/Qualifiers
     source         1..814
                    /organism="Bos taurus"
                    /mol_type="mRNA"
                    /db_xref="taxon:9913"
                    /clone="CSEQFXN41 testes"
                    /tissue_type="testes"
                    /note="Organ: testes; Vector: pBluescript SK+; Site_1:
                    NotI; Site_2: EcoRI; sequence 5' of the insert
                    (5'-NNN...NNNinsert)
                    GCGAATTGGACGTCCACGCGTGGCGGCGGGCGCTCGAG. Sequence 3' of
                    the inserts (AAGAATTCGATATCAAGCTTATCGATACCGTCGACCTCGAG.
                    normalized Rd 1 library, sequenced 3' with M13R primer."
BASE COUNT      254 a    161 c    173 g    226 t
ORIGIN

Query Match        21.3%;  Score 658.6;  DB 14;  Length 814;
Best Local Similarity   90.3%;  Pred. No. 1.6e-108;
Matches 758;  Conservative    0;  Mismatches  54;  Indels  27;  Gaps   4;

Qy       1348 ATGGTTCAAGCAGCAGATCTTCTTTCTCTGCCATTCATTCATTGCATCATGCATCCAG 1407
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        814 ATGGTTCAAGCAGCAGATCTTCTTTCTGCCATTCATTCATTGCATCATGTCATCCAG 755

Qy       1408 GCCCAGAATGAATTGACTACT-ACAAAAGGAGATCATCCAATTATGATGGGGTTTTGAACCCCTTGT 1466
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        754 GCCCAGAATGACACTAACAAAAGGAGACACCATCCCATTATGATGGGGTTTTGAGCCACTTGT 695
```

```
Qy   1467 GAACCAGAGGCTACTTCCACTTCTTCCCTGGATAATGCAAAAATAATTAAAAGGGAAGA 1526
          |||  || || |||||||||  | | |||||||||| |||||||||||||||||  | |||
Db    694 TAACCAAAAGACTACTTCCACCCACCTTCCCTGGATATGCAAAAATAATTAAAAGAGAGA  635

Qy   1527 AATGGTGAACTATTTGCAAGATTAATAGATAGAATAAAAACTGTCTGTGAGGTTGTGAA 1586
          ||||||| |||||||||||||| ||||| ||||||||||||||||||||||||| | |||
Db    634 AATGGTCAACTATTTGCAAGATGATAGAGATAGAATAAAAACTGTCTGTGAGGTGTGCAA  575

Qy   1587 TTTAACAAATTTACATTGTATCCTGGATTTTTTCTGTGAATTAGTAGTCAGTCACCATG 1646
          ||||||||||||||||||||||||||||||||||||||||||||| | | |  ||| ||
Db    574 TTTAACAAATTTACATTGTATCCTGGATTTTTTCTGTGAATTAGTGATCAATCACCTTG  515

Qy   1647 TGTTCTTTCAAGATCTCTTTACAAACCACTTTCTGTGGATAACAAAAGGTCTTTTGG 1706
          ||||||||| ||||||| || |||||||||||| ||| |||||||||||||||||||
Db    514 TGTTCTTTCGAGATCTATTTACAAACCACTTTCCTGGTGGATAACAAAAGGTCTTTTGG  455

Qy   1707 AACTCATCTCATGCCAAGAACAGTGAAAGATGGTGAAAGAGATGCACTTCGGTCTTTTGTTTGTCAGATCCTCCGA 1766
          |||||||||||||||||||||||||| |||||||||| ||||||||||||||| ||||| ||||||||||||||||
Db    454 AATCCATCCTCATGCAAGAGACAGTGGAAGATGGTGAAAGATGCTCTTCGGTCTTTTGTTTGTCAGATCCTCCGG-  396

Qy   1767 GTGCTTCTCCCCAAGTGCTACCTATAATAATAATCACCAGGGTAGGACTGTATGGACTCC 1826
          |||||||||||||||||||||| |||| |||||||||||||| ||| ||||||||||| |
Db    395 -TGCTCTCCCCAAGTGCTACCTATAATAATAATCACCAGGGTAGGACTGTATTGGACTCT  337

Qy   1827 TTTTGTTTACTCACTGTGTTGGGGCCATTCTGTAGTCTTTATTCAGATGCACATAACAGG 1886
          |||||| |||||| ||||||||||||||||||||||| |  | ||||| |||||||||||
Db    336 TTTTGTTTACTCGGTGTTGGGGCCATTCTGTAGTCTTTATTCAGGATCCTTAATCAACAGG  277

Qy   1887 GCTCGACGAGAGGATAAGCTTGGTCATAATCTTGTAGTCTTTATTCAGATTGCCACTTGCAGGGATGAG 1946
          ||||||||||||| || ||||||| |||||| ||| ||||||||||| ||||||||||||||||||||
Db    276 GCTCGGCAGAGAGGATAAACTCGGTCATATCTTGTAGTCTTCACTTGTACTTGCACCTTGCGAGGATG-  219

Qy   1947 TTTATGACATTTTATTTAATAGGCAGAGAGGTTGATGCAGGGAATTTTGTCACCATGTCGT 2006
          |||||||||||||| |  |  |      ||| |||||| ||| |||||| ||| || ||||
Db    218 ------------------------AGGCAGAGAAGGTTGATGCAGCAGCCCTTCATACTATGCTGT  180

Qy   2007 TGAAAACAGGAAACCCAAAAGGCAACAAGTTGGGCTTGTTAGGTACTGGGGTTCCTTTACCATA 2066
          ||||||||||||||| |||| || |||| |||||||||| ||||||||| |||||||||||||
Db    179 TGAAACAGGACCCCACAGGCCGCATTGGGCCTGTTTAGGTAACTGGCCTTTTACCATA  120

Qy   2067 ACCTTCGCATTATGATAGTACCTTCTAAGTGGCCTTGTTGGGAACTCTA-CAGTATG 2125
          ||||||||  ||||| || || ||||||||  ||||| |||| ||  | ||||| ||
Db    119 ACCTTCGAATTAATGATATCACGTACCTTCTAAGTGGCCTTCTCGAATTGGAGCTCTACCAGCATG  60

Qy   2126 CAGGATACTATTACATATATATGGTATCTCTGAATTCCTTTACGCATGGTTCATGTC 2184
          |||||||  |||| |||||||||||||||||||||||||||| |||||||| ||||| 1
Db     59 CACGAGTATTATTACATATATTGGTATCTCTCGAATTCTTTTATGCCTGGTTGCATGTC  1
```

RESULT 6

```
LOCUS       AK050805                3071 bp    mRNA            linear   HTC  05-DEC-2002
DEFINITION  Mus musculus 9 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:D030020M24 product:CORNEAL WOUND HEALING
            RELATED PROTEIN homolog [Rattus norvegicus], full insert sequence.
ACCESSION   AK050805
VERSION     AK050805.1  GI:26094130
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE   RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE   Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 3071)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Furuno,M., Hiramoto,T., Hiraoka,I., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M., and Hayashizaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)

```
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
```

```
FEATURES             Location/Qualifiers
     source          1..3071
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /clone="D030020M24"
                     /db_xref="FANTOM_DB:D030020M24"
                     /db_xref="taxon:10090"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /tissue_type="whole body"
                     /note="CORNEAL WOUND HEALING RELATED PROTEIN homolog
                     [Rattus norvegicus] (SPTR|Q9JI01), evidence: FASTY,
                     /dev_stage="9 days embryo"
     misc_feature    1..3071

BASE COUNT      811 a      611 c      686 g      963 t
ORIGIN

Query Match          21.1%;    Score 653.8;   DB 11;   Length 3071;
Best Local Similarity    88.1%;   Pred. No. 1e-107;
Matches  736;  Conservative    0;  Mismatches  92;  Indels   7;  Gaps   2;

Qy   1316  AGACCAGTGCTGTTGCAGAAGCTCAAAAAATTGATGGTTCAAGCAGCAGGAGATCTTCTTTCTG  1375
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    210  AGACCAGTGCTGTTGCAGAGCTCAGAAAACTGATGGTCCAGGCGGCGGAGGAGATCTTCTTTCTG  269

Qy   1376  CCATTCATATAATTCATTCATCATCATGGCATCCAGGCCCAGAATGATACTACAAAAAGGAGATC  1435
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    270  CCATTCACACCTCATTGCACCACGGCATCCAGGCCTCAGAATGGCACTACCAAAAGGAGAGACC  329

Qy   1436  ATCCAAATTATGATGATGGGGTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACCTACCTTCC  1495
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    330  ATCCAAATTATGATGGGGTTTTGAGCCCCTTGTTAACCAGAGGATTACTTCCACCCACCTTCC  389

Qy   1496  CTCGATATGCAAAAATAATTAATTAAAAGGGAAGAAAATGTGAACTATTTTGCAAGATTAATAG  1555
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    390  CTCGCTATGCAAAAATAATTAAAAGAGAGAAAATGGTCAACTATTTCTCAAGATTAATAG  449

Qy   1556  ATAGAATAAAAACTGTCTGTGTGGTTGTGAATTTAACAAATTTACATTGTATCCTGGATT  1615
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    450  ACAGATAAAAAACTGTCGTGTGAAGTCGTGAACTTACACTGTATCCTGGATT  509

Qy   1616  TTTTCTGTGAATTTAGTGAACAGTCACCATGGTGTTCTTCAGAGACTCTGTTACAAACCA  1675
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    510  TTTTCTGTGAATTTAGTGAGCAGTCACCATGGTGTTCTTCAAGATCTCTGTTACAAACCA  569

Qy   1676  CTTTTCTTCGTTGGTAGTAACAAAAAGGTTCTTTGGACTCATCTCAAGACACATGGTGAAAG  1735
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    570  CTTTTCCTTCGTTGGTAACAAAAAGGTCTTTGGCACTCATCTCATGCAAGACATGGTGAAAG  629

Qy   1736  ATGCACTTCGGGTCTTTGTCAGATCCTCCGAGTGTCTTCCCCAAGTGCTACCTATATAA  1795
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    630  ATGCTCGTGGCGTCCTTCGTCAG--TCCTCCGGGTGCTCTCCCCAAGTGTTGCCTAATAA  687

Qy   1796  TAATCACCAGGCTAAGGACTGTATCGGACTCCTTTGTTACTCACTGTGTTGGGCCATTCTG  1855
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    688  TAATCACCAGGCTAAGGACTGTATCGGACTCCTTTGTTACTCACTGTGTTGGGCCGTTTTG  747

Qy   1856  TAGTCTTATTCAGATCCATGGACAATAACAGGGCTCGACAGAGAGAGTAAGCTTGGTCATTAT  1915
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    748  TAGTCTTGTTGTTCAGATCCAATCACAGGGCTCGGCAGAGAGAGACAAGCTTGGTCACAT  807

Qy   1916  TCTTGAGGAATTGCCACCTTGGACGATGAGTTTATGACATTTTATTTAATAGGCAGAG  1975
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    808  TCTTGAGGAGTTTGCTACCTTGCAAGATCAGT-----AACAATTGTTTGACAGGGAGAG  862

Qy   1976  AAGGTTGATGCAGCGCTTCACCATGTCTGTTGTTGAAAACAGGAGAG  2035
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    863  AAGGTTGATGGCAGCGCTTCATACTTGCAAGAAGAG  922

Qy   2036  GCCTGTTTAGGTACCTGGGTCTCTTTACCATAAGCCATTATGATACAGTACAGTACCTTCTA  2095
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2036  GCCTGTTTAGGTACCTGGGTCTCTTTACCATAAGCCATTATGATACAGTACAGTACCTTCTA  2095
```

RESULT 7

```
LOCUS       BG547889                   768 bp    mRNA    linear   EST 04-APR-2001
DEFINITION  602576187F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4704212 5',
            mRNA sequence.
ACCESSION   BG547889
VERSION     BG547889.1   GI:13546554
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 768)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1543  row: n   column: 21
            High quality sequence stop: 738.
FEATURES             Location/Qualifiers
     source          1..768
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4704212"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_77"
                     /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
                     SfiI (ggcccctcggcc); Site 2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGCCGCCATATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGCGGGGCGGGCCGGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.9
                     kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH MGC Library."

BASE COUNT      241 a      152 c      179 g      196 t
ORIGIN

Query Match          21.1%;    Score 653.6;   DB 10;   Length 768;
Best Local Similarity    96.9%;   Pred. No. 1.3e-107;
Matches  698;  Conservative    0;  Mismatches  19;  Indels   3;  Gaps   3;
```

```
Db    923  GCCTGCTTAGGAACCTGGGTTCTTTACCATAAACCTCCGGATTATGATCCAGTATCTGCTC  982
           |||| ||||| ||||||||| |||||||||||||||||||||||| |||||||||||| ||
Qy   2096  AGTGGCTTTGAAATTGGAACTCTACAGTATGCAGGAGTACTATTACATATATTGGT  2150
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    983  AGTGGCTTTGAGCTGGAGCTGTACCAGCATGCATGAGTACTACTACTTACTGGT  1037

Qy   2149  GTATCTCCTGAATTCCTTTACGCATGGTTGATGTCAACATTGGTTGCCATGGCTC  2208
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GTATCTCTGAATTCCTTTACGCATGGTTGATGTCAACATTGAGTCGTGCCGATGGCTC  60

Qy   2209  TCAAATGGCAGGACGGAAAGGATAATGGAAGCAGCAGGAAAGGCCGTAGTAGTAAAAAAAC  2268
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  TCAAATGGCAGAGGGAAAGGATAATGGAAGCAGCAGAAAAGGCCGTAGTAGTAAAAAAAC  120

Qy   2269  AAAGAAAAAAGAAAAGAAAGTTCGCCCATTGAGCCGAGAGATCACAAATGAGCCAAGCATATCA  2328
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  AAAGAAAAAAAAGAAAAGTTCGCCCATTGAGCCGAGAGATCACAAATGAGCCTAGCATATCA  180

Qy   2329  GAACACATGTGTGGGAAGTGTTAAAACCATGGGGTAGGGATGGACGGGCAAAAGTACG  2388
```

```
Db   181 GAACAATGTGTGCTGGAAATGTTAAAACCATGGTAGCATTGACATGGACGGCAAAGTAG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2389 TAAACCGAAGTTTGAGCTTGATTAGTGAACAAGTTCGGTATGAACACAGGTTGCTTCCATT 2448
Qy  2389 TAAACCGAAGTTTGAGCTTGATTAGTGAACAAGTTCGGTATGAACACAGGTTGCTTCCATT 2448
Db   241 TAAACCGAAGTTTGAGCTTGATTAGTGAACAAGTTCGGTATGAACACAGGTTTGCTCCATT 300
Qy  2449 CAACAGTGTGATGACCCCGCGCCGCCCGAGCCCATCATGGTTATTGAGAAAATGTTCGACCT 2508
Db   301 CAACAGTGTGATGACCCCGCGCCGCCCGAGCCCATCATGGTTATTGAGAAAATGTTCGACCT 360
Qy  2509 CAATAA-ATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGCGACTAGTAAGCACT 2567
Db   361 CAATAACATATAGCCCTCCTCCTCAGTCTCCTGAACTGTATGTGCGACTAGTAAGCACT 420
Qy  2568 TTCAACAGGCAAAAATGATATTGAAAATATTCCTAACCCGACCATGAGGTTTAATAGAA 2627
Db   421 TTCAACAGGCAAAAATGATATTGAAAATATTCCTAACCCGACCATGAGGTTCATAGAA 480
Qy  2628 TTTTAAAGGTTGCCAAACCCAACTTTGTTGGTTATGAAGGTTATTGGCAGGAGGACACAAAA 2687
Db   481 TTTTAAAGGTTGCCAAGACCAACTTTGTGGTTTATGAAGGTTATTGGCAGGAGGACACAAAA 540
Qy  2688 AGGAATCTAAAGTTCCTCCCTGAATTTGATTTCTCTGCTCATTATTTCCCGTGTTGTGA 2747
Db   541 AGGAATCTAAAGTTCCTCTCGAATTTGATTTCTCTGCTCATTATTTCCCGTGTAGTGA 600
Qy  2748 AACTTGTTTTGAGAGAGACTGGGGAGGTGGGCTCAT-AAAGGGGGCAGAGTCTTCTTTTCAGACC 2806
Db   601 AACCTTGTTGAGAGAGACTGGGGAGGTGGGCCATGGCAGAGTCTTCTATCAGACC 660
Qy  2807 CAACTCTTAGAGGGGGGCCACCAACACCAGGCTCCAACATCACGGGAAGTGAGATGGAATTCCTTGG 2866
Db   661 CAATTCTTAGAGGGGGGCCATTCACGAGCCTCA-ATCACGGGAAGTGAGATGGAATTCCTTGGG 719
```

BASE COUNT   274 a   198 c   253 g   232 t
ORIGIN

```
                    NotI; Cloned unidirectionally. Primer: Oligo dT. Average
                    insert size 1.7 kb. Constructed by ResGen, Invitrogen
                    Corp. Note: this is a NIH_MGC library."
```

Query Match   20.9%; Score 648; DB 13; Length 957;
Best Local Similarity   87.7%;   Pred. No. 1.3e-106;
Matches 719; Conservative 0; Mismatches 100; Indels   1; Gaps   1;

```
Qy   541 GGAGGAAGTAGGCATAATAGTTATGTATGAAAGCTTCGTTAGATGATGACGATTAGGATGGGAG 600
             ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    26 GGCGCGGACGGCATCATGGTTATTGAGAAAGCTACTGTGAGACGACGATGCTTCGGGATGGGAG 85
Qy   601 CTCAGTATGCCAGAAAAAATGGAGAAAACGCAATACAAACTGGGGTGGGACATTACCAAGAT 660
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    86 CTCGGGGTCCCGGAAAAAATGGAGAAAAAGTAGCACAACAAGGTGGACATAACCAGGAT 145
Qy   661 TTTGAAGAAGCTTGTCGAGAATTAAAGTTGGGAGAAGCAATTCATGATAAGCTATTTGGT 720
             ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db   146 TTTGAAGAGTCCTTGTCGAGAGCTGAAGTTGGGAGAAACTGCCTTCATGATAAGCTATTTGGT 205
Qy   721 CTTTTTGAAGCCATGTCTGCTATTGAAATGGAATCCCAAGATGGATGCTGGCATGATT 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   206 CTTTTTGAAGCCATGTCTGCTATTGAAATGGAATGGAATCCCAAGAGTGGATGCCGGTATGATC 265
Qy   781 GGAAACCAAGTAATGAAAAAGTTCTCAAAATTTTGAAACAGCTATCAAGAATGGCACTATT 840
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   266 GGGGAACCAAGTGAAATAGAAAAGTTCTCAAATTTTGAACAAGCTATCAAGGATGGCACCATT 325
Qy   841 AAAATTAAAGATCTCACCTTGCCTGAACTGATGAGGTTTAGGGATTTATGGATACATGTTTTTCTGT 900
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   326 AAAATTAAAAGACCTCAGCCCTGACTGATGAGGATAATGACACCTGTTTCTGTGTGT 385
Qy   901 TTGATAACGTGGTTAGAAGGCCATTCACTGGACAGTATTTACGTGCCTTTACATT 960
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   386 TTGATCACATGGCTCGAAAGGCCATTCCTTCCTTGGCACAGACAGCAGTGTTTACGTGCCTTTACATT 445
Qy   961 CATAATCCCAGACTTTATATAGAGATTCCTGCTATGAGAAGGCCTTGCTCTGGGGATCTTGAAA 1020
             |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   446 CATAATCCCGACTTCATAGAGACAATCCTGCCAAGAAAGCTTTTGCTCTGGGGAAATCTTGAAG 505
Qy  1021 ATCTGTGACATTGCAAGGGAAAAGTAAAATAAAAGCTCGTCGTTTTTGAAGAGGAAGATTT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   506 ATCTGCGACAATGACGGGAAAAAGTAAAATAAAGCTCGTGTTTTCGAAGAGAAGGAAGATTTC 565
Qy  1081 CAGTCAATGACTTATGGATTTAAAATGGCTAACAGTGTGACAGATCTTCGAGTTAAGGC 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   566 CAGTCAATGACATACGGGATTTAAAATGGCCAACAGTGTGACAGATCTCCGAGTTACAGG 625
Qy  1141 ATGCTAAAAGTATGTGAGGATGACATGCAAAGAAGAGAGTAAAGTACTCGAGCAA 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   626 ATGCTAAAAGTGTGAGGATAATCTGAAAGGCCGAGTAAGAGTACTCGAGACAA 685
Qy  1201 GGAGGAAGAAGAGATCCAGGAGAGTGAACTAGAACACAACAAATGTTTAGCAGTATTCAGC 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   686 GGAGGAGAGAGATCCGAGGTCAACTAGAACACGGAGGTGCTTGCAGCATTCAGC 745
Qy  1261 AGAGTGAAATTTACTCGTGTTGTTACTGACAGTGCTTATAGCCTTTACTAGAAAAGAGACC 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   746 AGAGTGAAGTTCACGCGAGTGCTGCTCACAGTGGCTCATAGGCTTACTAAG-AAGAGAAC 804
Qy  1321 AGTGCTGTTGCAGAAGCTCAAAAAATTGATGGTTCAAGCAG 1360
             ||||||||||||||||||||||||||||||||||||||||
Db   805 AGTGCTGGTGCAAAAGGCTCAAAAACTGATGGGCCCAGCGG 844
```

```
VERSION     AI790514.1  GI:5338230
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished
COMMENT     Other ESTs: u102b03.y1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:994033
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 509.
FEATURES             Location/Qualifiers
     source          1..828
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2064845"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Sugano mouse kidney mkia"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site 1: DraIII
                     (CACTGTG); Site 2: DraIII (CACCAGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCCTTTTTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a DraIII adaptor [TGTTCGGCCTTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCAGTG). XhoI should
                     be used to isolate the cDNA insert. Size selection was
                     performed to exclude fragments <1.5kb. Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science). Custom primers for
                     sequencing: 5' end primer CTTCTGCCTTAAAAGCTGGG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      208 a    197 c    177 g    240 t      6 others
ORIGIN

Query Match     20.7%;  Score 641.4; DB 9;  Length 828;
Best Local Similarity   86.4%;
  Matches  716;  Conservative    0;  Mismatches  112;  Indels    1;  Gaps    1;

Qy    673 TGTTCGAGAATTAAAGTTGGGAGAACTACTTCATGATAAGCTATTTGGTCTTTTTTGAAGCC 732
          ||||||||| || ||||||||| ||||||| |||||| |||||||| ||| ||||||||||||
Db    828 TGTTCGAGAGCTGAAGTTGGAAAGACTACTGCTTCATTGATAAGCTGTTTGTTCTTTTGAAGCC 769

Qy    733 ATGTCTCTGCTATTGAAAATGATGGATGACCCAGATAGGGATGGCTGGCATGAAAACCAAGTT 792
          ||||||| ||||||||| |||||| || || || ||| ||||||| | |||||||||||||||
Db    768 ATGTCTCCTATTGAAAATGATGATTCATTAAGATGGGAGGCGGGGTATGATCNGAACCAAAGTG 709

Qy    793 AATCGAAAAGTTCTCAATTTGAACAAGCTATCAAGGATGGCCACTATTAAAAATTAAAGAT 852
          ||| | |||||||||||||| || ||||||||||||||||||  ||||||| |||||||||
Db    708 AATAGNAAAGTTCTCAATTTTGANCAAGCTATCAAGGATGGCAC-ATTAAAATTAAAGAC 650

Qy    853 CTCACCTTGCCTGAACCTGATAGGGATTATGGATACACATGTTTTGCTGTTTGATTAACGTGG 912
          |||| || || ||||||| ||||| |  ||||| ||| ||||| || ||||||||  ||| ||
Db    649 CTCAGCTGCTGCNTGAACTCATAGGAATAATGGACACCTGTTTCTGCGTTTTGATCACATGG 590

RESULT 10
BM979511/c
LOCUS       BM979511                616 bp    mRNA     linear   EST 21-FEB-2003
DEFINITION  UI-CF-DU1-adr-h-18-0-UI.s1 UI-CF-DU1 Homo sapiens cDNA clone
            UI-CF-DU1-adr-h-18-0-UI 3', mRNA sequence.
ACCESSION   BM979511
VERSION     BM979511.1  GI:19600035
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 616)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
   MEDLINE   97044377
   PUBMED    8889548
COMMENT     Contact: McCray, PB
            McCray Lab
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M.J. Welsh, University of Iowa
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research

Qy    913 TTAGAAAGGCCATTCACTGGCACAGACAGACAGTATTTACGTTGCTTTGCCTTTACATTCATTCATTCAGAC 972
          ||||||||||||||| | |||||||||||| |||| |||||| |||||| |||||||||||  |  ||||||
Db    589 CTCGAAAGGCCATTCCTTGGCACAGACAGCAGTGTTTACGTGCCTTTACATTCATAATCCCGAC 530

Qy    973 TTTATAGAAGATCCTGCTATGAAAAGGCTTTTGCTTGCTTGGGGAATTCTTGAAAAATCTGACATT 1032
          || ||||||||||||||  ||||||| ||||| |||| | |||||||||||||||| |||| ||||
Db    529 TTCATAGAAGATCCTGCCATGAAAAGCTTTTTGCTCTGGGGAATTCTTGAAAATCTGCGACATT 470

Qy   1033 GCAAGGGAAAAAGTAAAATAAAACCTGCTGTCTTTTTTGAAGAGAGAAGAGATTTTCAGTCAATGACT 1092
          ||| ||||||||| |||||||||| ||||| ||||| ||||||||| |  | ||||||||||||||| |
Db    469 GCACGGGAAAAAAGTAAAATAAAGCTGCTGTTTTTCGAAGAGAGGAAAGGATTTTCAGTCAATGACA 410

Qy   1093 TATGGATTTTAAAAATGGCTAACAGTGTGACAGATCTTCCGAGTTACAGGCATGCTAAAAGAT 1152
          || |||||||||||||||| |||||||||||||||| | ||||||||||  |||||||||| |
Db    409 TACGGATTTTAAAAATGGCCAACAGTGTGACAGATCTCCGAGTTACAGGATGCTAAAAGGAT 350

Qy   1153 GTGGAGGGATGACATGCAAAGAAGGAAGAAGAAGTAGTAGTCGAAGTCGACAAGAGAAGAAAGA 1212
          ||||| | ||||| | ||||| | ||| |  ||  ||| || ||||||||||||||| |||| |
Db    349 GTGGAAGAGATGATCGCAAAAGGCGGAGAGTAAAGAGAGTACTCGAAGTCGACAAGAGAGAGAGA 290

Qy   1213 GATCCAGAAGTTGACTAGAACACCAACAATGTTTAGCAGTGTATTCAGCAGCAGAGTGAAATTT 1272
          |||||||| ||||| |||||| |||| ||||| | | ||| ||| ||||||||| ||||| ||
Db    289 GATCCAGAGGTCGAACTAGAACCCCAGACAATGTCTTGGCAGCCATCAGCAGCAGTGAAGTTC 230

Qy   1273 ACTCGTGTGTTTACTGACAGTGGCTTATAGCCTTTACTAAGAAAAGAGACCAGTGGCTGTTGCA 1332
          ||  | ||| | || ||||  ||||| ||||||| || ||||||||| ||| | ||| ||||||
Db    229 ACGGCAGTGGCTGCTCACCAGTCGCTCATAGCCTTCATGGAGAAAAGAAGTGCTCGTTGCA 170

Qy   1333 GAAGCTTCAAAAAATTGATGGTTCAAGCAGCAGACAGATCTTCTTTCTGCCATTCATAATTCATTG 1392
          || ||| ||||||  |||||||||| |||||||||  |||||||||||||||||||  | | ||||
Db    169 GAGGCTCAGAAAACTGATGGTTCAGGCAGCAGACCTTCTTTCTGCCATTCACACCTCATTG 110

Qy   1393 CATCAATGGCATCCAGCCCCAGAATGATTACTACAAAAGGAGAATCATCCAATTATGATGGGT 1452
          || | | |||||| |||  | |||||  | | | ||| ||   ||||||||||| | ||||||
Db    109 CACCACGGCATCAAGCGCTCGAAATGGCCGAAATGGACCATCCAATTAAAATGGGT 50

Qy   1453 TTTGAACCCCTTGTGAACCCAGAGGCTACTTCCACCTACTTCCCTCGAT 1501
          |||||  ||||||||||| |||||  |||||||| |  ||||||||| |
Db     49 TTTGAGCCCCTTGTTAACCAGAGATTACTTCCCCCCGCTTCCCTCGNT 1
```

Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES   Location/Qualifiers
     source   1..616
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-DU1-adr-h-18-0-UI"
                     /tissue_type="Primary Lung Epithelial Cells"
                     /dev_stage="Adult"
                     /lab_host="DH10B (life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-CF-DU1"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-DU1 is a normalized cDNA library containing the
                     following tissue(s): Primary Lung Epithelial Cells The
                     library was constructed according to Bonaldo, Lennon and
                     Soares, Genome Research, 6:791-806, 1996. First strand
                     cDNA synthesis was primed with an oligo-dT primer
                     containing a Not I site. Double stranded cDNA was ligated
                     to an EcoR I adaptor, digested with Not I, and cloned
                     directionally into pT7T3-Pac vector. The oligonucleotide
                     used to prime the synthesis of first-strand cDNA contains
                     a library tag sequence that is located between the Not I
                     site and the (dT)18 tail. The sequence tag for this
                     library is GGCTGTAGGC.

BASE COUNT   190 a   123 c   98 g   205 t
ORIGIN

TAG_LIB=UI-CF-DU1
TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
TAG_SEQ=GGCTGTAGGC"

Query Match   19.7%;   Score 610;   DB 12;   Length 616;
Best Local Similarity   100.0%;   Pred. No. 9.6e-100;
Matches 610; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   2487 AGTTCAAGGAAATGTCTGACCTCAATAATAATAATAGCCCTCCTCCTCAGTCTCCTGAACTGT 2546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   616 AGTTCAAGGAAATGTCTGACCTCAATAATAATAGCCCTCCTCCTCAGTCTCCTGAACTGT 557

Qy   2547 ACGTGGCAGCTAGTAAGCACTTTCAACAGGCAAAAATGATATATGGAAAAATATTCCTAACC 2606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   556 ATGTGGCAGCTAGTAAGCACTTTCAACAGGCAAAAATGATATATGGAAAAATATTCCTAACC 497

Qy   2607 CGGACCATGAGGGTTAATAGAATTTTAAAGGTTGCCAAACCCAACTTTGTGGTTTATGAAGT 2666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   496 CGGACCATGAGGGTTAATAGAATTTTAAAGGTTGCCAAACCCAACTTTGTGGTTTATGAAGT 437

Qy   2667 TATTGGCAGGAGGACAACAAAAAGGAAATCTAAAGTTCCTCCTGAATTTGATTTCTCTGCTC 2726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   436 TATTGGCAGGAGGACAACAAAAAGGAAATCTAAAGTTCCTCCTGAAATTTGATTTCTCTGCTC 377

Qy   2727 ATAAAATATTTTCCTGTTGTTGTTTGTTTTGAGAGAGACTGGGAGGTGGGCCATAAAAGGG 2786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   376 ATAAAATATTTTCCTGTTGTGAAAACTTGTTTTGAGAGAGACTGGGAGGTGGGCCATAAAAGGG 317

Qy   2787 GCAGAGTCTTCTTTCAGACCCAACTCTTAGAGGGCACATCACCCAGGGCTCCCAACATCACGGG 2846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   316 GCAGAGTCTTCTTTCAGACCCAACTCTTAGAGGGCACATCACCCAGGGCTCCCAACATCACGGG 257

Qy   2847 AAGTGAGATGGATTCTTCTTGGGTAACACACTCATTATAGGAATACTTTTAGTTTGACAGCC 2906
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   256 AAGTGAGATGGATTCTTCTTGGGTAACACACTCATTATAGGAATACTTTTAGTTTGACAGCC 197

Qy   2907 TTATATGACATGAAGATGAATGAAAAAACTGCTGTTTTAAAGTGGTTATTATGTTCCATGGAAGAAA 2966
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   196 TTATATGACATGAAGATGAATGAAAAAACTGCTGTTTTAAAGTGGTTATTATGTTCCATGGAAGAAA 137

Qy   2967 CTGGTCTTATTGAATGCAATGAATGAGAACGTTATATATGTTTATTTTAATTACAGAATTAATCACAA 3026
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   136 CTGGTCTTATTGAATGCAATTGATGAAGCGTTATATATGTTTATTTTAATTACAGATTTAATCACAAA 77

Qy   3027 TCATTTTTATGAATGATTGAGTTGAGTGAAAATAGTGTTTATAAAGGTAATAAAATTTTCTTGAC 3086
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   76 TCATTTTTTATGAATGATTGAGTGAGTGAAAATAGTGTTTATAAAGGTTAATAAAATTTTCTTGAC 17

Qy   3087 AAAAAAAAAA 3096
          ||||||||||
Db   16 AAAAAAAAAA 7

RESULT 11
CA977646
LOCUS   CA977646   900 bp   mRNA   linear   EST 06-JAN-2003
DEFINITION AGENCOURT1 11295472 NIH_MGC_164 Mus musculus cDNA clone
            IMAGE:30146905 5', mRNA sequence.
ACCESSION   CA977646
VERSION   CA977646.1 GI:27510300
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 900)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe and Dr. Mina
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM0063 row: f column: 02
            High quality sequence stop: 555.

FEATURES   Location/Qualifiers
     source   1..900
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:30146905"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_164"
                     /note="Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2:
                     NciI; Non-normalized full-length enriched library from
                     pooled mouse embryonic limb, maxilla and mandible, day
                     10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
                     Cloned directionally; priming method: Oligo-dT.
                     enrichment: >1k bp, Average insert size 1.8k bp. Priming
                     sequence: 5'GACTAGTTCTAGATCGCGAGCGGCCGCCC(T) 3'. Tissue
                     contributed by David Rowe. Library constructed by ResGen,
                     Invitrogen Corp."

BASE COUNT   235 a   213 c   212 g   239 t
ORIGIN

                                                                    1 others

Query Match   19.4%;   Score 599.2;   DB 14;   Length 900;
Best Local Similarity   85.5%;   Pred. No. 8.1e-98;
Matches 700; Conservative   0; Mismatches   94; Indels   25; Gaps   2;

Qy   1535 ACTATTTTGCAAGATTAATAGAATAGAAATAAAAACTGTCTGTGAGGTTGTGAAATTAACAA 1594
          ||||||||| ||||||||||||||||||| ||| ||||||||||||||||||||||||||
Db   3 ACTATTTTCTTCCAAGATTAATAGAACAGAATAAAAAACTGTCTGTGAAGTCGGAAGTTAACAA 62

Qy   1595 ATTTTACATTGTATCCTGGATTTTTTCTGTGAATTTAGTGAACAGTCACCATGTGTTTCTTT 1654
          || |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 ACTTACACTGTATCCTGGATTTTTCTGTGAATTTAGTGAGACAGTCACCATGTGTTCTTT 122

Qy   1655 CAAGATCTCTGTTACAAACCACACTTTCCTGGGTGGGATAACAAAAAGGTCTTTGGAACTCATC 1714
          |||||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||||
Db   123 CAAGATCTCTGTTACAAACCACTTTCCCTCGTGGGATAACAAAAAGGTCTTTGGCACTCATC 182

RESULT 12
BI602827
LOCUS   BI602827   697 bp   mRNA   linear   EST 07-SEP-2001
DEFINITION   603247187F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5294804 5',
             mRNA sequence.
ACCESSION   BI602827
VERSION   BI602827.1 GI:15495766
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 697)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgaps-r@mail.nih.gov
             Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
             Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM11745 row: n   column: 21
             High quality sequence stop: 643.

FEATURES   Location/Qualifiers
     source   1..697
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5294804"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTVN-3',
                     size-selected for average insert size 2.3 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation). Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT   208 a   116 c   185 g   188 t
ORIGIN

Query Match   19.3%; Score 598.8; DB 12; Length 697;
Best Local Similarity   99.5%; Pred. No. 9.9e-98;
Matches 611; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy   1715 TCATGCAAGACATGGTGAAAGATGCGACTTCGGTCTCTTTGTTCAGATCCTCGAGTGCTTTC 1774
Db   183 TCATGCAAGACATGGTGAAAGATGCTCTGCGGTCCTTCGGTCGCTCTC 240

Qy   1775 CCCCAAGTGCTGCCTATATAATAATCACCAGAAGCTAAGGACTGTATGGACTCCTTTGTTAC 1834
Db   241 CCCCAAGTTGCCTATATAATAATCACCAGAAGCTAAGGACTGTATCGACTCCTTTGTTAC 300

Qy   1835 TCACTGTGTTCGGCCATTCTGTAGTCTTATTCAGATCCATGGACAATAACAGGGCTCGACA 1894
Db   301 TCACTGTGTTCGGCCGTTTGTTAGTCTTGTTCAGATCCATGCCAATAACAGGGCTCGGCA 360

Qy   1895 GAGAGATAAGCTTGGTCATATTCTTGAGGAATTTGCCACCTTGCAGGATGAGTTTATGAC 1954
Db   361 GAGAGACAAGCTTGGTCACATTCTTGAGGAGTTTGCTACCTTGCAAGATG 410

Qy   1955 ATTTTATTTTAATTAGGGAGGAGGAGGAGGTTGAAGCACGGCTTCACACCATGCTGTTGAAACAG 2014
Db   411 ---------AGGCAGAGAGGTTGATGCAGCCGGCCTTCATACTATGTTGTTGAAACAG 457

Qy   2015 GAACCCCAAAGGCAAACATTTGGGCCTTTAGGTACCTGGGGTTCCTTTACCATAACCTTGC 2074
Db   458 GAGCCTTCAGAGACAACAATCTGGCCTTGCTTAGGAACCTGGGTTCTTTACCATAACCTCCG 517

Qy   2075 ATTATACAGTACCTTCTAAGTGGCTTTGAATTGGAACTTCTACAGTATCACGGAGTAC 2134
Db   518 ATTATAATCCAGTATCTGCTCAGTGGCTTTGAGTGTTGACATCGCATGAGTAC 577

Qy   2135 TATTACATATATTGGTATCTCTCTGAATTCCTTTACGCATGGTTGAGTGTCAACATTGAGT 2194
Db   578 TACTACATCTACTGGTACCTCCCGAGTTCCGTGTAGTCATGGCTGCTCAACCTCGAGC 637

Qy   2195 CGTGCCGATGGGTCTCAAAATGGGAGGAGGAGAAAAGGAGATAATGAAGAGACGCAGCAGCACGAAAAGGGCGT 2254
Db   638 CGCGGCCTGAGGGCCTCTCAGATGGCAGAGGGATCATGGAGGAGGAGGATCACAGATTAAGAGGCCGC 697

Qy   2255 AGTAGTAAAAAAACAAAGAAAAAAAAAGAGTTCGCCCCAGAGAGATCACAATG 2314
Db   698 AGCCGCCAAAAAACCAAGAAAGAAAAAAGTTCGTCCGTTGAGCCGAGAGATCACAATG 757

Qy   2315 AGCCAAGACATATCAGAGAACATGTGTGCTGGAAGTGTTTAAA 2353
Db   758 AGCCAGGCCTATCAGAGAACATGTGTGCTGGGATGTTCAA 796

Qy   459 GAGAGGGGAGGAGGGGCGGGGCGGCGGGGCGGCCCAGGCGCCGCGTCGTTATTTCCGTGGTCCGGACAGTG 518
Db   6 GAGAGGGGAGGAGGGGGCGGGGGCGGGCGGGCCGAGGCGCCGCGCTTGTTATTTCCGTGGTCCGGACAGTG 65

Qy   519 CGTGGCGGCGGCGGCGGGCGGTGACCACGGGGGAAGGAAGTAAGGCCATAATAATGGTTATGAAAGCTTCTGTAGA 578
Db   66 CGTGGCGGCGGCGGGGGGTGACCACGGGGGAGAAGGAAGTAAGGCCATAATAATGGTTATGAAAGCTTCTGTAGA 125

Qy   579 TGATGACGAATTCAGGAGGAGCTCAGTATGCCCAGAAAAAAATGGGAGAAGCAATACAAA 638
Db   126 TGATGACGATTCAGGATGGGAGCTCAGTATGCCAGAAAAAAATGGGAGAAGCAATACAAA 185

Qy   639 CTGGGTGGACATTACCCAAGATTTTGAAGAAGCTTGTGTCGAGAATTAAAGTTGGGAGAACT 698
Db   186 CTGGGTGGACATTACCCAGATTTTGAAGAAGCTTGTTGTCGAGAATTAAAGTTGGGAGAACT 245

Qy   699 ACTTCATGATAAGCTATTTGGTCTCTTTTGAAGCCCATGTCTGCTATTGAAAATGATGGATCC 758
Db   246 ACTTCATGATAAGCTATTTGGTCTCTTTTTGAAGCCCATGTCTGCTATTGAAAATGATGGATCC 305

Qy   759 CAAGATGGATGCTGGCATCATGATTGGAAAACCAAGTTAATCGAAAAGTTCTCAATTTGAACA 818
Db   306 CAAGATGGATGCTGGCATCGATTGGAAAACCAAGTTAATCGAAAAGTTTCTCAATTTGAACA 365

Qy   819 AGCTATCAAGGAATGGCACTATTAAAATAAAATTAAAGATCTCACCTTGCCTGAACTGATAGGGGAT 878
Db   366 AGCTATCAAGGAATGGCACTATTAAAATAAAATTAAAGATCTCACCTTGCCTGAACTGATAGGGGAT 425

Qy   879 TATGGATACATGTTTTTGCTGTTTGAATAACGTGGTTAGAAGGCCATTCACTGGCACAGAC 938
Db   426 TATGGATACATGTTTTTGCTGTTTGGAATAACGTGGTTAGAAGGCCATTCACTGGCACAGAC 485

Qy   939 AGTATTTTACGTGCCTTTACATTCATAATCCAGACTTTATAGAAGATCCTGTCATGGAAGGC 998
Db   486 AGTATTTTACGTGCCTTTACATTCATAATCCAGACTTTATAGAAGATCCTGTCATGGAAGGC 545

Qy   999 TTTTGCTTCTGGGGAATCTTGAAAATCTGTAGACATTGCAAGGAGAAAAAGTA-AATAAAGCTG 1057
Db   546 TTTTGCTTCTGGGGAATCTTGAAAATCTGTAGACATTGCAAGGAGAAAAAGTACAATAAAGCTG 605

Qy   1058 CTGTCTTTGAAGAG 1071
Db   606 CTGTCTCTGAAGAG 619

RESULT 13
AW007583/c
LOCUS   AW007583   598 bp   mRNA   linear   EST 08-MAR-2000

```
DEFINITION  wt02f11.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2506317 3',
            mRNA sequence.
ACCESSION   AW007583
VERSION     AW007583.1  GI:5856446
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.
            , Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arraying: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 715   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1..598
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2506317"
                     /sex="pooled"
                     /tissue_type="colon"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Co3"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
                     was prepared from 12 pooled bulk tumor samples and primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library went through one round of
                     normalization."
BASE COUNT      189 a    122 c     98 g    188 t      1 others
ORIGIN
```

```
Query Match    19.3%;  Score 597;  DB 9;  Length 598;
Best Local Similarity  99.8%;
Matches  597;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy     2490 TCAAGGAAAATGTCTGACCTCAAATAAAATAAGCCTCCTCCTCCTCGAACTGTATG 2549
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      598 TCAAGGAAAATGTCTGACCTCAAATAAAATAAGCCTCCTCCTCCTCGAACTGTATG 539

Qy     2550 TGGCAGCTAGTAAGCACTTTCAACAGGCAAAAATGATATTGGAAAATATTCCTAACCCGG 2609
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      538 TGGCAGCTAGTAAGCACTTTCAACAGGCAAAAATGATATTGGAAAATATTCCTAACCCG 479

Qy     2610 ACCATGAGGTTAATAGAAATTTAAAGGTTGCCAAAACCCAACTTTGTTGGTTATGAAGTTAT 2669
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      478 ACCATGAGGTTAATAGAAATTTAAAGGTTGCCAAAACCCAACTTTGTTGGTTATGAAGTTAT 419

Qy     2670 TGGCAGGAGGACACAAAAAAGGAATCTAAAGTTCCTCCTGAAATTGAATTTCTGCTCATA 2729
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      418 TGGCAGGAGGACACAAAAAGGAATCTAAAGTTCCTCCTGAAATTGAATTTCTGCTCATA 359

Qy     2730 AATATTTTTCCTGTGTTGAAACTTGTTTGAGAGAGACTCGGGAGGTGGCCATAAAGGGGCA 2789
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      358 AATATTTTTCCTGTGTTGAAACTTGTTTGAGAGAGACTCGGGAGGTGGCCATAAAGGGGCA 299

Qy     2790 GAGTCTTCTTTCAGACCCAACTCTTAGAGGGGCACCACATCACCAGGCTCCACATCACGGGAAG 2849
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1
```

```
Db      298 GAGTCTTCTTTCAGACCCAACTCTTAGAGGGGCACCACATCACCAGGCTCCACATCACGGGAAG 239
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     2850 TGAGATGGGGATTTCTTGGGTAACAACTCCATATATAAGGAATAAGGAATACTTTAGTTGACAGCTTA 2909
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      238 TGAGATGGGGATTTCTTGGGTAACAACTCCATATATAAGGAATAAGGAATACTTTAGTTGACAGCTTA 179

Qy     2910 TATGACATGAATGAAAACTGCTGTTTTATTAATTATTATGTTCCATGAAGAAACTG 2969
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      178 TATGACATGAATGAAAACTGCTGTTTTATTAATTATTATGTTCCATGAAGAAACTG 119

Qy     2970 GTCTTATTGAATGCAATGAAACGTTATATACGGTTTATTACAGATTTAATCACAAATCA 3029
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      118 GTCTTATTGAATGCAATGAAACGTTATATACGGTTTATTACAGATTTAATCACAAATCA 59

Qy     3030 TTTTTTATGAATGATTGAGTGAATAAAATAGTGTTTTTATAAAGGTTAATAAAATTTCTTGACA 3087
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       58 TTTTTTATGAATGATTGAGTGAATAAAATAGTGTTTTTATAAAGGTTAATAAAATTTCTTGACA 1
```

```
RESULT 14
LOCUS       BU058931                 718 bp    mRNA    linear   EST 26-AUG-2002
DEFINITION  UI-M-FR0-cam-e-22-0-UI.r1 NIH_BMAP_FR0 Mus musculus cDNA clone
            IMAGE:6414333 5', mRNA sequence.
ACCESSION   BU058931
VERSION     BU058931.1  GI:22499220
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 718)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jim Lin, University of Iowa
            Site 2: Not I; The library was constructed according
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..718
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6414333"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /clone_lib="NIH_BMAP_FR0"
                     /note="Organ: Brain; Vector: pYX- Asc; Site 1: EcoR I;
                     Site 2: Not I. The library was constructed according
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured RNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with oligo-dT
                     primer containing a Not I site. Double strand cDNA was
                     size selected according to mRNA size fraction, ligated
                     with EcoR I adaptor, digested with NotI and then cloned
                     directionally into pYX-Asc vector. The library tag
                     sequence located between the EcoR I site and the polyA tail
                     is AGCAGACAG. This library was created for the University
                     Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                     Developing Mouse Nervous System' supported by National
                     Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
```

```
                          program coordinator."
BASE COUNT     222 a     159 c     165 g     171 t
ORIGIN

Query Match                                        1 others
Best Local Similarity   89.1%;   Score 593.8; DB 13;   Length 718;
Matches 640; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy    896 GCTGTTTGATAAGGTGTGGTTAGAAGGCACAGCCATTCACTGGACACACAGTAATTACTGCCTTT   955
          |||||||||| |  |||||| || ||||||||| |||| |  ||||| | ||||| ||||||||||
Db      1 GCTGTTTGATCACATGGCTGCGAAGGCCATTCCTTGGCACACAGCACAGTGTTTACTGCCTTT    60

Qy    956 ACATTCATTAATCCAGACTTTATAGAAGGATCCTGCTATGAAGGCTTTTGCTCTGGGAATCT   1015
          |||||||||||     ||| ||||||||||||| |||||||||||||||||| |||||||||
Db     61 ACATTCATAATCCCGACTTCATAGAACATGAAGCTTTTGCTCTGGGAATCT              120

Qy   1016 TGAAAATCTGTGACATTCAAGGGAAAAAGTAAATAAAAGCTGGTGTTTGAAGAGGAAG   1075
          |||||| || ||  || ||||||  ||||||||  |||| |||| |||| ||||||||
Db    121 TGAAGATCTGCGACATTCCAGCGGGAAAAAGTAAATAAAGCTCGTGTTTTCGAAGAGGAAG   180

Qy   1076 ATTTTCAGTCAATGACTTATGGAATTAAAATGGCTAACAGTGTGACAGATCTTCGAGTTA   1135
          |||||||||||||||| |||||||||  ||||| ||||||||||||||| ||||||||||
Db    181 ATTTTCAGTCAATGACATACGGAATTAAAATGGCCAACACAGTGTGACAGATCTCCGAGTTA   240

Qy   1136 CAGGCATGCTAAAAGATGTGGAGGATGACATGCAAAGAAGAGACTAAAGAGTACTCGAAGTC   1195
          |||||||| | ||| |||||||||| ||  ||||||| ||||||||||| |||| ||||||
Db    241 CAGGGATGCTAAAAGGATGTGGAACAGTGAAAGCGAGTGAAGAAGTACTCGAAGTC        300

Qy   1196 GACAAGAGGAGAGAAAAGAGGATCCAGAAGTGAACTRAGAACACCAACAAATGTTTAGCAGTAT   1255
          ||||||||||||||| || |||||||||| ||||  || ||||||||| || ||||||||||||
Db    301 GACAAGGGAGAGAGGATCCAGAGGTCGAACTAGAACACCAGCGCTTGGCAGCAT          360

Qy   1256 TCAGCAGAGTGAAATTTACTCGTGTGTTACTGACAGTGTCTTATAGCCTTTACTTAAGAAAG   1315
          |||||||||||||||| |||| |||| ||||| ||||| ||| |||||||| ||||||||||
Db    361 TCAGCAGAGTGAAGTTCACGCGAGTGCTCCTCACAGTGCTCATAGCCTTCATCAGTTAAGAAAG   420

Qy   1316 AGACCAGTGCTGCTTGTTGCAGAAGCTCATCAAAAATTGATGTTCATCGCAGCAGAGATCTTCTTTCTG   1375
          |||||||| |||||||| ||||||| |||||| ||  |||| ||||||||| ||||| |||||||||||
Db    421 AGACCAGTGCTGCTTGCAGAGGCTCAGAAACTGATCGTCCAGGCGGCGAGACCTTCTTTCTG        480

Qy   1376 CCATTCATAATTCATTGCGATCATGGCCATCCAGGCCCCAGAATGATCACTACCAAAAGGAGATC   1435
          |||||||| |||| ||||||||||||||||||||| ||||| |||| || |||||||||||||||
Db    481 CCATTCACACCTCATTGCGACCATCCGGCCATCCAGGCTCAGATTGGTAACTACCAAAAGGAGACC   540

Qy   1436 ATCCAATTATGATGGGGTTTTGAACCCCCTTGTGAACCAGAGGCTACTCTTCCACCTACCTTCC   1495
          ||||||| ||||| |||||||||||||||| ||||||||| ||| ||||||||| || ||||||
Db    541 ATCCAATTATGATGGGGTTTTGAGCCCTTGGAACAGAGGATTACTTCACCCACCTTCC          600

Qy   1496 CTCGAATGCAAAAATAATAATAAAAGGAAGAAAATGGTGAACTATTTTGCAAGATTAATAG   1555
          |||||||||||| |||| |||||  || ||||||||||||  ||||||||||||| |||||
Db    601 CTCGCTATGCAAAAATAATAAAAGGGTCAACTACTATTTCTCAAGATTAATAG              660

Qy   1556 ATAGAATAAAAAACTGTCTGTGAGGTTGTTGAATTAACAAATTTAACTTACATTGTATCCTGGA   1613
          ||||||| |||||||| ||||||| |||| ||||| ||||| |||| |||||||||||||||||
Db    661 ACAGAATAAAAAACTGTCTGTGAAGTTCGTTGAACTTACCAAAACTTACACTGTATCCTGGA   718

RESULT 15
BG112812
LOCUS       BG112812                 972 bp    mRNA       linear   EST 30-JAN-2001
DEFINITION  602285005F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4372320 5',
            mRNA sequence.
ACCESSION   BG112812
VERSION     BG112812.1  GI:12606318
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 972)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
```

```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1032  row: j  column: 01
            Clone_id: LLAM10032
FEATURES             Location/Qualifiers
     source          1..972
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MAGE:4372320"
                     /tissue_type="osteosarcoma, cell line"
                     /clone_lib="NIH_MGC_86"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: bone; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.533 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
                     High quality sequence stop: 658.
BASE COUNT     285 a     232 c     214 g     241 t
ORIGIN

Query Match
Best Local Similarity   96.6%;   Score 586; DB 10;   Length 972;
Matches 663; Conservative 0; Mismatches 15; Indels 8; Gaps 6;

Qy   2090 CTTCTAAGTGGGCTTTTGAAATGGGAACTTCACAAGGTTGGATGTGAACAGTTGCACGAGTACTATTACACATATTATTGG   2149
          ||||||||||||||||||||||||||| || |||||||||||||||||||||||||||||||||||||| |||||||||
Db      3 CTTCTAAGTGGGCTTTTGAA-TGGAACTTCACTACACATGCACGAGTAGTATTTACATATA-TGG          60

Qy   2150 TATCTCTCTGGAATTCCTTTAGGCAAGGGTTGAGTGTCACAGTTGAGTCGTCGTGCCGATGGCTCT   2209
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TATCTCTCTGGAATTCCTTTAGGCAATGGTTGAGTGTCAACATTGAGTCGTGCCGATGGCTCT         120

Qy   2210 CAAATGGCAGAGGAGAAAGGAGGATAAATGGAAGAGACAGCAGAGAAAGGCCGTAGTAGTGAAAAAACA   2269
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||| |||||||||
Db    121 CAAATGGCAGGAGAAAGGAGGATAAATGGAAGAGCAGCAGAGAAAGGCCGTAGTAGTAAAAAAACA       180

Qy   2270 AAGGAGAAAAGAGAAAGTTCGCCCATTGAGCCGAGAGATCAC-AATGAGCAAGCAGCATATCAG   2329
          ||||||||||||||||||||| ||||||||||||||||||||  |||||||||||||||||||
Db    181 AAGGAAAAAGAAAAGTTCCCCCATTGAGCCGAGAGATCACAAATGAGCAAGCAGCATATCAG       240

Qy   2330 AACATGTGTGCTGGAATGTTTAAAACCATGTGTAGCATTTGACATGGGCAAAAGTACGT   2389
          |||||||||||||||||||||||| ||||||| ||||| |||||| || ||||||||||
Db    241 ACAGTGTGCTGGAATGTTTAAAACCCATTGAGCATTTGACATGGGCAAAAGTACGT         300

Qy   2390 AAACCGAAGTTTGAGCTTGAATAGTGAACAAGTTCGGTATGAACAAGTTCGTTGCTCCATTC   2449
          |||||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||
Db    301 AAACCGAAGTTTGAGCTTGATAGTGAACAAGTTCGGTATGAACAAGTTCGTTGCTCCATTC       360

Qy   2450 AACAGTGTGATGACCCCGCGCGCCAGTGCACTACTTACAGTTCAAGGAAAATGTCTGACCTC   2509
          |||||||||||||||||||||||||||||||||||| |||||||| |||||||||||||||
Db    361 AACAGTGTGATGACCCCGCCGGCCAGTGCACTACTTACAGTTCAAGGAAATGTCTCGACCTC     420

Qy   2510 AATAAAATATAGCCCTCCCTCAGTCTCCTCAGTCTCTGAACTGTATGTGGCAGCTAGGAGCTTT   2569
          |||| ||||||||||||||| ||| ||||||||||||||||||||||||||||||||||||||
Db    421 AATAAAATAGCCCTCCCTCCTCAGTCTCCTCAGCTGTATGTGGCAGCTAGGAGCACTTT        480

Qy   2570 CAACAGGCCAAAAATGATATTGGAAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATT   2629
          |||||||||||||||||| |||||||||| |||||||||||||||||||||||||||||||
Db    481 CAACAGGCCAAAAATGATATTGGAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATT   540

Qy   2630 TTAAAGGTTGCCAAACCCAAACTTTGTGGTT-ATGAAGTTAATTGGCAGGAGGGAGGACAHAAA   2687
          |||||||||||||||| || ||| |||||||  ||||||| ||||||||||||||||| |||
Db    541 TTAAAGGTTGCCAAACCCAACTTTGTGGTTCATGAAGTTAATTGGCAGGAGGGAGGACAACAA   600

Qy   2688 AGGAATCTAAAGTTCCTCCCTGAATTGATTTCTCTGCTCATAAAATATTTCCTGTGTGA   2747
          |||||||||||||||| ||||||||||||| ||||||| |||||||||||||| |||||
Db    601 TTAAAGGTTGCAAACCCAACTTTGTGGTTCATGAAGTTATTGGCAGGAGGACACAACAA   660
```

```
                    ||||||||||||||||||||||| |||||||||||||||||||||||          || ||| |      |||||
Db   601 AGGAATCTAAAGTTCCTCCTGAA-TTGATTTCTCTGCTCATA---ATATTTCCTGTGTGA 656

Qy  2748 AACTTGTTTGAGAGAGACTGGGGAGG 2773
              ||||| |||||||| |||| || |
Db   657 AACTGTTTGAGAGAGGACTGGGGAGGTGG 682
```

Search completed: November 25, 2003, 02:50:11
Job time : 6201 secs

OM nucleic - protein search, using frame_plus_n2p model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: November 24, 2003, 22:01:10 ; Search time 96.5 Seconds
(without alignments)
10184.818 Million cell updates/sec

Title: US-10-001-857-42
Perfect score: 5439
Sequence: 1 ttcctcacgaaactcccagg..............atttcttgacaaaaaaaaaa 3096

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues
Listing first 45 summaries

Total number of hits satisfying chosen parameters: 2215726

Post-processing: Minimum Match 10%
Maximum Match 10%

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2l/USPTO_spool/US10001857/runat_24112003_162854_9527/app_query.fasta_1.3271
-DB=A_Geneseq_19Jun03 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -CUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US1001857@CGN_1_1_0 @runat_24112003_162854_9527 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -LONGLOG
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7 -XGAPEXT=0.5 -FGAPOP=6

Database :
A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3796	69.8	733	23	ABP52902	Human lung specifi
2	3732.5	68.6	725	23	AAO19400	Human molecule for
3	1298.0		784	23	ABB60388	Drosophila melanog
4	1237	22.7	239	23	ABG69816	Human secretory pr
5	392	7.1	108	22	AAU32414	Novel human secret
6	158.5	2.9	681	22	ABG93650	Herbicidally activ
7	153.5	2.8	156	22	ABG14149	Novel human diagno
8	153	2.8	708	23	ABB91504	Herbicidally activ
9	146.5	2.7	120	22	ABU52794	Human transmembran
10	146.5	2.7	574	22	AAM52322	WASP homolog prote
11	146.5	2.7	574	22	AAG67370	Amino acid sequenc
12	146	2.7	416	23	ABG79529	Human tumour suppr
13	145.5	2.7	694	23	ABB92421	Herbicidally activ
14	145	2.8	149	22	AAO05346	Human polypeptide
15	145	2.6	559	22	AAM52317	Human Scarl Protei
16	145	2.6	559	22	AAG67345	Amino acid sequenc
17	145	2.6	559	22	AAB20494	Human Pablo (pro-a
18	145	2.6	559	22	AAB73399	Human WAVE1 (WAVE/
19	144	2.6	1134	23	AAE24341	Human lung-specifi
20	143.5	2.6	19938	24	ABP76682	Streptomyces virid
21	143	2.6	598	22	ABG14000	Novel human diagno
22	143	2.6	780	24	ABU00297	Human novel polype
23	143	2.6	1248	20	AAY13464	Human diaphanous p
24	143	2.6	19938	24	ABB98398	Streptomyces virid
25	143	2.6	551	23	ABG93105	Human diaphanous p
26	141.5	2.6	560	22	ABG21040	Novel human diagno
27	141.5	2.6	5877	22	AAE34702	S. cerevisiae BAX-
28	141	2.6	191	21	AAG29836	Novel human mucin
29	141	2.6	199	21	AAG29835	Human mucin (MUC-1
30	140.5	2.6	19938	24	ABP76680	Arabidopsis thalia
31	140	2.6	541	22	AAW37148	Arabidopsis thalia
32	140	2.6	783	19	AAW37151	Streptomyces virid
33	140	2.6	787	19	AAW37152	Mammalian Ena (Men
34	140	2.5	802	19	AAW37153	Mouse neural Mena+
35	140	2.5	802	22	AAU09139	Mouse neural Mena+
36	139.5	2.5	1243	24	AAE34865	Mouse neural Mena+
37	139.5	2.5	1243	24	AAB71620	Mammalian enabled
38	139	2.5	214	17	AAR86913	Human kinases and
39	139	2.5	250	21	AAY67470	Human WNK4 protein
40	138.5	2.5	641	21	AAY67469	Cotton fiber-speci
41	138.5	2.5	641	21	AAY82327	Np70 protein carbo
42	138.5	2.5	641	22	AAY72165	Human Npw38BP1 tra
43	138.5	2.5	641	22	AAB47514	Np70 protein seque
44	138.5	2.5	647	21	AAB53462	Human RNA metaboli
45	138.5	2.5	903	23	ABP65235	NpwBP; Homo sapie

ALIGNMENTS

RESULT 1
ABP52902
ID   ABP52902 standard; Protein; 733 AA.
AC   ABP52902;
DT   05-NOV-2002 (first entry)
DE   Human lung specific protein sequence SEQ ID NO:145.
XX
KW   Human; lung; lung specific nucleic acid; LSNA; lung specific protein;
KW   LSP; cytostatic; gene therapy; vaccine; metastasis; lung cancer;
KW   squamous cell carcinoma.
OS   Homo sapiens.
XX
PN   WO2020264788-A2.
XX

```
PD   22-AUG-2002.
XX
PF   20-NOV-2001; 2001WO; 2001WO-US45080.
XX
PR   20-NOV-2000; 2000US; 2000US-252054P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Macina RA, Recipon H, Chen S, Sun Y, Liu C;
DR   WPI; 2002-657601/70.
XX
PT   New lung specific nucleic acid useful in gene therapy or in vaccines
PT   for treating lung cancer (e.g. squamous cell carcinoma) or
PT   non-cancerous lung diseases, as well as for diagnosing, monitoring or
PT   staging these diseases -
XX
PS   Claim 11; Page 239-242; 282pp; English.
XX
CC   The present invention describes an isolated lung specific nucleic acid
CC   (LSNA), comprising a sequence that: (a) encodes any of the 93 amino acid
CC   sequences comprising 17 - 733 amino acids, given in ABP52873 to ABP52965;
CC   (b) comprises any of 115 sequences comprising 148 - 3193 base pairs (bp),
CC   given in ABQ75262 to ABQ75376; (c) selectively hybridises to (a) or (b);
CC   or (d) has 60 % sequence identity to (a) or (b). LSNA and lung specific
CC   protein (LSP) sequences have cytostatic activity and can be used in gene
CC   therapy and vaccines. LSNAs and LSPs are useful for diagnosing and
CC   monitoring the presence and metastases of lung cancer in a patient. An
CC   antibody that specifically binds to an LSP can be used for determining
CC   the presence of an LSP in a sample, as well as for treating a patient
CC   with lung cancer, particularly by inducing an immune response against
CC   the lung cancer cell expressing the LSNAs or LSPs. In particular, these
CC   LSNAs and LSPs are useful for identifying, diagnosing, monitoring,
CC   staging, imaging and treating lung cancer (e.g. squamous cell carcinoma)
CC   and non-cancerous disease states in lung.
XX
SQ   Sequence   733 AA;
```

US-10-001-857-42 (1-3096) x ABP52902 (1-733)

```
Alignment Scores:

Pred. No.:                0
Score:              3796.00
Percent Similarity:  99.59%
Percent Identity:    99.59%
Best Local Similarity: 99.79%
Query Match:             23

                            Length:          733
                            Matches:         732
                            Conservative:      0
                            Mismatches:        1
                            Indels:            3
                            Gaps:              0
                            DB:                0
```

```
Qy    556  ATGGTTATGAAACCTTCTGTAGATGATGAGGATTCAGGATGGGAGCTTCAGTATGCCAGAA   615
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MetValMetLysAlaSerValAspAspAspSerGlyTrpGluLeuSerMetProGlu      20

Qy    616  AAAATGGAGAAAAGCAATACAAACTGGGGTGGACAATTACCACCAAGACATTTGAAGAAGCTTGT   675
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21  LysMetGluLysSerAsnThrAsnTrpValAspIleThrGlnAspPheGluGluAlaCys    40

Qy    676  CGAGAATTAAAGTTGGGAGAACTACTTCATCGTAAGCTATATTGGTCTTTTTGAAGCCATG   735
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41  ArgGluLeuLysLeuGlyGluLeuLeuHisAspLysLeuPheGlyLeuPheGluAlaMet    60

Qy    736  TCTGCTATTGAAATGATGGATCCCAAGATGCCTGGGCATGATTGAAAACCAAGTTAAT     795
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  SerAlaIleGluMetMetAspProLysMetAspGlyMetIleGlyValAsnAsn         80

Qy    796  CGAAAAGTTCTCAATTTGAACAAGCTATCAAGGATGGCACTATTATAAAATTAAAGATCTC    855
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81  ArgLysValLeuAsnPheGluGlnAlaIleLysAspGlyThrIleLeuLysLysLysAsp   100

Qy    856  ACCTTGCCTGAACTGATAGGGGATTAAGGAACTACTTCATGTTTTCGTGTTCATAAGCTGGTTA   915
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101  ThrLeuProGluGluIleGlyIleMetAspThrCysPheCysLeuIleThrTrpLeu      120

Qy    916  GAAGGCCATTCACTGGCACAGACAGTATTTACGTGCCTTACAATTCATAATTCAGAGACTTT   975
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  GluGlyHisSerLeuAlaGlnThrValPheThrCysLeuLeuTyrIleHisAsnProAspPhe  140

Qy    976  ATAGAAGATCCTGCTATGAGAAGATGCTTTGCCTGGGAATCTGAAAAATCTGACATTGCA    1035
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141  IleGluAspProAlaMetLysAlaLeuGlyIleLeuGluLysLysSerAspIleAla      160

Qy   1036  AGGGAAAAAGTAAATAAAAAGCTTGCTGTTGTTTTCAAGGAAGATTTCAGTCAAGACTTAT   1095
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161  ArgGluLysValAsnLysLysAlaAlaValPheGluValGluAspPheSerMetThrTyr   180

Qy   1096  GGAATTAAAATGGCTAACAGTGTGACAGATCTTCGAGTGAGGTTACAGGCATGCTAAAGATGTG  1155
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  GlyPheLysMetAlaAsnSerValThrAspLeuArgValThrGlyMetLeuLysAspVal   200

Qy   1156  GAGGATGACATGCAAAAGAAGAGTAAAGAAGAGTACTCGAAGTCGACAAGGAGAAAGAGAT   1215
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201  GluAspAspMetGlnArgArgValSerThrArgSerArgGlnGlyLysGluArgAsp      220

Qy   1216  CCAGAAGTTGAACTAGAACACCAACAATGTTTAGCAGTAGTTCAGCAGAGTGAAATTTACT   1275
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221  ProGluValGluLeuGluHisGlnGlnCysLeuAlaValPheSerArgValLysPheThr   240

Qy   1276  CGTGTGTTACTGACAGTGCTTATAGCCCTTACTAAGGAAGAAGAAGACCAGTGCTGCTTGTTGCAGAA  1335
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  ArgValLeuLeuThrValLeuIleAlaPheThrLysGluGluGluThrSerAlaAlaGlu   260

Qy   1336  GCTCAAAAATTGATGTGTTCAAGCAGCAGAGATCTTCTTTCTGCCATTCATAATTCATAATGCAT  1395
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261  AlaGlnLysLeuMetValGlnAlaAlaAspLeuLeuSerAlaIleHisAsnSerLeuHis   280

Qy   1396  CATGGCATCCAGGCCCAAGATGATACTACAAAAGGAGGATCATCCAATTATGATGGTTTT    1455
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281  HisGlyIleGlnAlaGlnAspAspThrThrLysGlyGlySerSerAsnTyrAspGlyPhe   300

Qy   1456  GAACCCCTTGTTGAACCAGAGGCTACTTCCACTTCCCTCGATATGCCAAAAATAATT       1515
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  GluProLeuValAsnGlnArgLeuLeuProProThrPheProArgTyrAlaLysSerIleIle  320

Qy   1516  AAAAGGGGAAGAAATGGTGAACTATTTTGCAAGATTAATAGAATAAAAACTGTCTGT       1575
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321  LysArgGluGluMetValAsnTyrPheAlaArgLeuIleAspArgIleLysThrValCys   340

Qy   1576  GAGGTTGTGAATTTAACAAATTAACATTGTATCCCTGGATTTTTCTGTGGAATTTAGTGAA   1635
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    341  GluValValAsnLeuThrAsnLeuHisCysIleLeuLeuAspPhePheCysGlyPheSerGlu  360

Qy   1636  CAGTCACCATGTGTTCTTCAAGATCCTCGTTCAAACCACACTTCCTGGTGGATAACAAA     1695
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  GlnSerProCysValLeuSerArgSerLeuValGlnThrThrPheLeuValAspAsnLys   380

Qy   1696  AAGGTCTTTGGAACTCATCTCATGCAAGACACATGGTAAGACTATGGGTTCTTTTGTC     1755
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381  LysValPheGlyThrHisLeuMetGlnAspThrMetValLysSerAlaLeuArgSerPheVal  400

Qy   1756  AGATCCTCGGAGTGCTGTCCCCAAGTGCTACCTATATAATCACCAGGCTAAGGACT       1815
           |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    401  -AspProPro-ValLeuSerProLysCysTyrLeuTyrAsnAsnHisGlnAlaLysAspSer  420

Qy   1816  GTATGGACTTCCTTTGTTACTTCACTGTGTTCGGCCATTTGTTGCTTATTCAGATCCATG    1875
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  ValTrpThrSerPheValThrSerLeuValArgProPheValAlaTyrSerAspProCys   440

Qy   1876  GACATAACAGGGCTCGACAGAGATAAGCTTGGTCATAATTCATTACTTCATAATTCTCCACCT  1935
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441  AspIleThrGlyLeuAspArgAspLysLeuGlyHisAsnSerLeuLeuHisAsnSerProPro  460

Qy   1936  TGCAGGATGAGTTATGACATTTTATTTAATAGGCA-GAGAAGGTTGATGCAGCCGCTTC     1994
           |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    461  CysArgMetSerTyrAspIleLeuPheAsnArgAla-GluGluValAspAlaAlaAlaAlaLeuH  480

Qy   1995  ACACCATGCTGTTGAAACAGGAAACCCAAAGCCAACATTTGCTGTTTAGGTACCTGGGG     2054
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    480  isThrMetLeuLeuLysGlnProGlnArgGlnHisLeuValCysLeuGlyThrTrpV
Qy    2055 TCCTTTACCATAACCTTCGCATTATGATTGATACAGTACACCTTCTTAAGTGGCTTGAATTGGAAC 2114
Db    500  alLeuTyrHisAsnLeuArgIleMetIleGlnTyrLeuLeuSerGlyPheGluLeuGluL 520
Qy    2115 TCTACAGTATGCACGAGTACTATTACATTATAATATTATTGGTATTCTTCTCTGAATTCTTACGGCAT 2174
Db    520  euTyrSerMetHisGlyTyrTyrIleTyrTrpTyrLeuSerGluProPheLeuTyrAla T 540
Qy    2175 GGTTGATGTCAACATTGAGTCTGTCGCCGATGGCTCTCAAATGCGCAGGAAAGGAATAATGG 2234
Db    540  rpLeuMetSerThrLeuSerArgGlaAAspGlySerGlnMetAlaGlyGlnIleMetG 560
Qy    2235 AAGAGGAGCAGAAAGGCGTAGTAGTAAAAACAAAGAAAAAGAAAGTTCGCGCCAT 2294
Db    560  luGluGlnGlnLysGlyArgSerLysLysThrLysLysLysLysValArgProL 580
Qy    2295 TGAGCCGAGAGATCACAAATGAGCCAAGCATATCAGAACATGTGTGCTGGAAATGTTTTAAA 2354
Db    580  euSerArgGluIleThrMetSerGlnAlaTyrGlnAsnMetCysAlaGlyMetPheLysT 600
Qy    2355 CCATGGTAGCATTTGACATGGACGGGCAAAGTACGTAACTACCGGAAGTTGAGCTTGATAGTG 2414
Db    600  hrMetValAlaPheAspMetAspGlyLysValArgLysProLysPheGluLeuAspSerG 620
Qy    2415 AACACAGTTCGTATGAACACAAGTTTGTCCATTCAACAGTTGTGATGACCCGCGCCGCCAG 2474
Db    620  luGlnValArgTyrGluHisArgPheAlaProPheAsnSerValMetThrProProProV 640
Qy    2475 TGCACTACTTACAGTTCAAGGAAATGTCTGACCCTCAATAAATATAGCCCTCCTCCTCCTCCAGT 2534
Db    640  alHisTyrLeuGlnPheLysGluMetSerAspLeuAsnLysTyrSerProProProGlnS 660
Qy    2535 CTCCTGAACTGTATGTGGCAGCTAGTAAGCACTTTCACACGGCAAAAATGATATTGGAAA 2594
Db    660  erProGluLeuTyrValAlaAlaSerLysHisPheGlnThrAlaLysMetIleLeuGluA 680
Qy    2595 ATATTCCTAACCGGACCATGAGGTTATTAGAGATTTTAAAGGTTCCCAAACCAACTTTG 2654
Db    680  snIleProAsnProAspHisGluValAlaArgIleLeuGluLysValAlaLysProAsnPheV 700
Qy    2655 TGGTTATGAAGTTATTCGGCAGAGGACACAAAAAGGAATCTAAAGTTCCTCCTGAATTG 2714
Db    700  alValMetLysLeuLeuAlaGlyHisLysSerLysValProProGluPheA 720
Qy    2715 ATTTCTCTGCTCATAAAATATTTTCCTGTTGTTGAAACTTGTT 2755
Db    720  spPheSerAlaHisLysTyrPheProValValLysLeuVal 733
```

RESULT 2

```
AAO19400
ID    AAO19400 standard; Protein; 725 AA.
XX
AC    AAO19400;
XX
DT    10-DEC-2002  (first entry)
XX
DE    Human molecule for disease detection and treatment protein #3.
XX
KW    Human; molecule for disease detection and treatment; MDDT; gene therapy;
KW    cytostatic; antiarteriosclerotic; hepatotropic; anti-HIV; antiallergic;
KW    antiinflammatory; antiasthmatic; cerebroprotective; nootropic;
KW    neuroprotective; antiparkinsonian; cardiant; antianginal.
XX
OS    Homo sapiens.
XX
XX
PN    WO200270709-A2.
XX
PD    12-SEP-2002.
XX
PF    08-FEB-2002; 2002WO-US03709.
XX
PR    09-FEB-2001; 2001US-268117P.
PR    15-FEB-2001; 2001US-269618P.
PR    23-FEB-2001; 2001US-271118P.
PR    07-MAR-2001; 2001US-274486P.
PR    09-MAR-2001; 2001US-274436P.
PR    28-NOV-2001; 2001US-334209P.
PR    01-FEB-2002; 2002US-353284P.
XX
PA    (INCY-) INCYTE GENOMICS INC.
XX
PI    Lal PG, Baughn MR, Yao MG, Walla NK, Elliot VS, Xu Y;
PI    Honchell CD, Yue H, Ding L, Gietzen KJ, Ison CH, Lu DAM;
PI    Hafalia AJA, Ghandi AR, Thangavelu K, Sanjanwala MM, Tang YT;
PI    Ramkumar J, Griffin JA, Swarnaker A, Azimzai Y, Sapperstein SK;
PI    Burford N, Lee EA, Lu Y, Tran UK, Marquis JP;
XX
DR    WPI; 2002-713453/77.
DR    N-PSDB; AAL49929.
XX
PT    New human molecules for disease detection and treatment (MDDT), useful
PT    for diagnosing, treating and preventing diseases or conditions
PT    associated with the aberrant MDDT expression, e.g. cancer, AIDS,
PT    asthma, diabetes, hepatitis -
XX
PS    Claim 1; Page 128-130; 177pp; English.
XX
CC    The present invention relates to human proteins and coding sequences of
CC    molecules for disease detection and treatment MDDT. The sequences can be
CC    used in the treatment of diseases associated with the decreased
CC    expression or overexpression of MDDT, such as cell proliferative (cancer,
CC    atherosclerosis, hepatitis), autoimmune/inflammatory (e.g. AIDS,
CC    allergies, Addison's disease, asthma), developmental (dwarfism, renal
CC    tubular acidosis), neurological (e.g. stroke, Parkinson's disease,
CC    epilepsy) and cardiovascular (congestive heart failure, myocardial
CC    infarction, angina pectoris) disorders. The present sequence is a protein
CC    of the invention.
XX
SQ    Sequence  725 AA;
```

```
US-10-001-857-42 (1-3096) x AAO19400 (1-725)

Alignment Scores:
Pred. No.:           1
Score:               3732.50
Percent Similarity:  98.23%        Matches:       725
Best Local Similarity: 98.23%      Conservative:  723
Query Match:         68.62%        Mismatches:    0
                                   Indels:        2
                                   Gaps:          1
DB:                  23
```

```
Qy    556  ATGGTTATGAAAGTTCTCGTAGATGTATGATGACGATTCAGGATGGGAGCTCAGTATGCCAGAA 615
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1    MetValMetLysAlaSerValAspAspAspAspSerGlyTrpGluLeuSerMetProGlu 20

Qy    616  AAAAATGGAGAAAAGCAATACAAACTGGGTGGACAATACCCAAGAATTTTGAAGAAGCTTGT 675
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    21   LysMetGluLysSerAsnThrArgTrpValAspIleThrGlnAsnPheGluGluAlaCys 40

Qy    676  CGAGAATTAAAAGTTGGGAGAAACTACTTCATGATAAGCTATTTGGTCTTTTTGAAGCCATG 735
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    41   ArgGluLeuLysGluGlyLeuLeuHisAspLysPheGlyLeuPheGluAlaMet 60

Qy    736  TCTGCTATTGAAATGATGGATCCCAAGATGGAATGCTGGCATGATGAAACCAAGTTAAT 795
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61   SerAlaIleGluMetMetAspProLysMetAspAlaGlyMetIleGlyAsnGlnValAsn 80

Qy    796  CGAAAAGTTCTCAATTTTGAAACAAGCTACTCAAGGAGGCACTATTAAAATTAAAGATCTC 855
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    81   ArgLysValLeuAsnPheGluGlnAlaIleAlaIleLysAspSerGlyThrIleLysIle 100

Qy    856  ACCTTGCCTGAACTGATAGGGATTATGGATACATGCATATGCTGTTGTTGATAAGTGTTA 915
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101  ThrLeuProGluLeuIleGlyIleMetAspThrCysPheCysCysLeuIleThrTrpLeu 120
```

```
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL04491.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 7956; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signaling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL16176-ABL30511), and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   784 AA;
```

```
US-10-001-857-42 (1-3096) x ABB60388 (1-784)

Alignment Scores:
Pred. No.:                     2.07e-114
Score:                         1298.50
Percent Similarity:            55.42%
Best Local Similarity:         38.11%
Query Similarity:              23.87%
DB:                            22

Length:      784
Matches:     306
Conservative: 139
Mismatches:  265
Indels:      93
Gaps:        18
```

```
Qy    472  GCGGGCGGCGGCCGAGCGGCGTGCTTA------TTTCCGTCGTCCGGACAGTGCGTGGCG  525
           ||||||||||||||  | | |   ||||      ||||||    |||||||||||||||||
Db     34  AlaAlaAlaAlaArgAlaAlaSerSerLeuGlnAspGluTrpSer--------------  48

Qy    526  GCGCGGGGTGACCACGGGAGAAGTAGGCATAATG------------------GTTATG  564
           || |||||| | |     |     |||||||||||                   |||||
Db     49  ----------ValGlyGlyPheLeuAspProAspValGlnArgThrMetArg  64

Qy    565  AAAGCTTCGTGATGATGACGATTCAGGATGGGAGCTCAGTGTATGCCAGAAAAATGGAG  624
           ||||    ||| |   ||    ||| |||||||
Db     65  SerGlySerLeuAlaGluAspSerIleThrGlnTyrProMetHis--------  78

Qy    625  AAAAGCAATACAAACTGGGTGGACATTACCCAAGATTTTGAAGAGAGCTTGTCGAGAATTA  684
           |||||    |    ||   |||    ||| |||  ||||||||
Db     79  --------GlyTrpValThrLysGluProHisSAspAlaCysAlaGluLeu  94

Qy    685  AAGTTGGGAGAACTACTTCATGATAAGCTATTTGGGTCTTTTTGAACCATGTCTGCTATT  744
           ||||||  | | ||  ||||  |||   |||   ||||  |  |  | |||||||||  |
Db     95  GlnProGlyGluLeuLeuGlnAspMetLeuPheGlyLeuPheGluGluValAlaAlaIle  114

Qy    745  GAAAATGATGGATCCCAAGATGGATGCTGGCATGATTGGAAACCAAGTTAATCGAAAAGTT  804
           ||||  ||  ||| |    |||||| |  ||  | | ||   | ||||| ||  ||||
Db    115  GluIleMetAspProLysMetGlyMetGlyHisAspLeuGlyThrProSer  134

Qy    805  CTCAATTTTGAACAAGGTATCAAGGAATGGCACTATTAAAATTAAAGATCTCACCTTGCCT  864
           ||||||||||||  || || ||||||||| ||  | |||||||||   ||| ||| ||||
Db    135  ProSerPheGluAlaAlaIleLysIleAlaThrGlyAlaIleLeuLysAspLeuThrProSer  154

Qy    865  GAACTGATAGGGATTATGGATACACATGTTTTTGCGTTTGATAACGTGGTTAGAAGGCCAT  924
           |||| || | ||  || ||| |  ||||| |||  ||||  | |  ||||||| ||  |||
Db    155  GluLeuIleGlyIleTyrAspAlaLeuPheSerCysLeuValSerIleTrpLeuGluGlyAsn  174

Qy    925  TCACTGGCACAGACACTATTACGTGCCTTTACATTCAATTCCAGACTTTATATAGAGAT  984
           |||::|||   | |    ||| |  |   ||:::|||  ||:::||| | ||:::|||
Db    175  SerMetAspGlnValLeuPheThrCysLeuTyrLeuHisAlaProAlaGlnIleLysAsp  194

Qy    985  CCTGCTATGAAAGGCTTTTTGCTCTGGGAAATCTGGGAATTCAGTCATGACTTATGA  1044
           ||||:||||||||||||||:||:||||| | | |||||| |||||||||||| |||
Db    195  LysAlaLeuArgValPheCysThrAlaAlaArgLeuIleValValIleLysIle  214

Qy   1045  GTAAATAAAGCTGCTCGTTTTTGAAGAGGAAGAGATTTTCAGTCATCATGACTTATGGA  1098
           |:::| ||||| |||  | || ||||||||||||| :::|||||:|:|:||| | |||
Db    215  IleAlaValAlaAlaAlaAsnGluGluGluPheGln-----LeuTyrGlyAsnSer  232

Qy   1099  --TTTAAAATGGTTCTTAACAGTGTGACAGATCTTCAGACGATCTGCAGTGCTAAAAGATG  1155
             ::|||||||||| | ||| ||||  | |||||||||||| ||| |||| |  || | |
Db    233  AlaLeuLeuAlaLaGluLysAlaGlnProAlaThrValTyrSerSerLeuLysAspVal  252

Qy   1156  GAGGATGACATGCAAAGAAGAAGAGTAAAGAAGTACTCGCAAGGAGAAGAAAAGAGAT  1215
           |||||:|  |||  |    |  ||||||||||   |||||| |||||| | || ||
Db    253  GluAspGluLeuLeuGlyLysCysLysLysLeuThrSerThr------  266

Qy   1216  CCAGAAGTTGAACTAGGAACACCAACAATGTTTAGCGATTCGAGACGTGAAATTTACT  1275
           ||||||||||| | |||   | ||  | ::||||||||  |||||||:::||:|:|||
Db    267  --------GluAspTrpMetAlaValValHisArgLeuArgPheMet  279

Qy   1276  CCGTGTGTTACTGACAGTGCTTATAGCCTTTACTAAGAAGGAAGACCAGTGCTGTTGCAGA  1335
           ||:||||| |||  |||| ||| |:|||| ||||  |||||||  | |  ||  |:|||
Db    280  ArgHisLeuPheGlnValIleTyrHisValGluGlnMetAlaSerAsnAspThrValAsp  299

Qy   1336  GCTCAA-----------AAATTGATGGTTCAAGCACCAGATGTTCTTCTGCCATTCAT  1383
           || |||           ||| :||||||| |  | |||  | |||||||||||  |
Db    300  AspLysValAspProIleTyrLysIleLeuLeuValAlaSerGluMetLeuProGlyIleArg  319

Qy   1384  AATTCAATTCCATCATCANGGCATCCAGGCCCAGAAATGTGAACTATTTTGCAGATTAAATGATAGA  1440
           ||||||||||||||||| ||  |:|| |||  ||| || | |||:|||:|  |||| ::|
Db    320  AsnThrLeuAspArgValGlyThrArgGlu--------LysGlySerAspAlaPro  335

Qy   1441  ATTAATGAATGGGTTTTGAACCCCTTGTGAACCAGAGGCTACTTCCACTTCCTCGA  1500
           ::|||   ||||| || |||  |||||||  || ||||||||||||||||||:|||
Db    336  AsnProMetGlyPheSerProArgIleHisAspArgSerGlnProProAlaPheProArg  355

Qy   1501  TATGCAAAATAATTAAAAAGGGAGAAAATGGTGAACTATTTTGCAAGATTAATAGGATAGA  1560
           ||||| ||:|| ||||||||||||| | ||||||   ||| |||||| ||   |||||||
Db    356  SerIleLeuArgAspArgProSerTyrGlnPheLeuGluGlnMetIleSerArg  375

Qy   1561  ATAAAAACGTCTGTGAGTTGTGAATTTAACAAAATTACAATGTATCCTGGATTTTTTC  1620
           ||||||| |||||| |||| |||||||||||||| |  ||:||||  |:|||:|||||
Db    376  PheLysTyrAlaCysLysLysValThrLysTyrTyrSerAlaLeuAsnPhePhe  395

Qy   1621  TGTGAATTTAGTGAACAGTCA---CCATGTGTTTCTTCAAGATCTCTCGTTACAAACCACT  1677
           ||:||||||:||||||||||    ||||||||| |:|:|||||||||:::|||||||||||
Db    396  IleGluTyrSerLysSerGlyGlnCysIleLeuGluSerArgSerValLeuGlnThrLeu  415

Qy   1678  TTCCTGGTGGATAACAAAAAGGTCTTTGGAACTCATCTCATGCAAGACATGGTGAAAGAT  1737
           ||:|||||||||| | ||||||||||||  ||:|| |:|||| ||| ||||||:|||||
Db    416  PheSerAlaAsnMetAlaHisGlyLysMetAlaHisGlyIleLysLeuProMetLysGlyArgHis  435

Qy   1738  GCACTTCGGGTCTTTTGTCAGATCCTCCGAGTGCTCTTCCCCAAGTGCTTACCTATATAATA  1797
           || |||  |||||| |||||||||||  |||||||   ||||||||| | || |||||  |
Db    436  ValGlnValPhe-AsnSerProPro-ValLeuAsnAlaLysHisProValAlaAlaA  455

Qy   1798  ATCACCAGGCTAAGGACGCTATCGACTCCTTTGTTACTCACGTGTT------CGGCCAT  1851
           ||||||||  || |||| |||||| |:::|||:|||| | ||:|||      ||| |||
Db    455  spProLysValGlyGlnHisLeuGluAsnPheArgTyrCysIleAsnMetAsnThrP  475

Qy   1852  TCCTGAGTCTTATTCAGATCCATGGACAATAACAGGGGCTGCAGAGAGATAAGCTTGGTC  1911
           |||||||:||:| |||||||||  |  | |  | |||||||| ||||||| || |||||
Db    475  heThrGlnPheIleArgIleCysPheAsnArgAlaArgGlnArgAspLysLeuGlyLeuAlaAla  495

Qy   1912  ATATTCTTGAGAATTTCCCACCTTGCAG---------GATGAGTTTATGA  1953
           ::|||:|||||||||||||||||||||          ||||||:|||||
Db    495  rgLeuLeuGluAsnPheAspThrIleGlnValAspAlaAlaArgLeuAspSerMetMetA  515

Qy   1954  CATTTTATTTTAATAGGCGGAGA--------GAAGGTTGATGAGGCCTTCACCAGTGTT  2007
           ||||||||||||||||||| ||         |||:||:|||||:|||||:|||||:|||
Db    515  snGlnLeuAlaAsnGlu-ArgAlaMetGluGly-AsnGluProMetAlaThrAlaLeu--  533
```

```
Qy  2008 GAAAACAGGAACCCGAAAGGCAACATTTGGGCCTGTTGTTAGGTACCTGGGTCCCTTTACCATAA 2067
                                       |||||||||||||||||||
Db   534 ------------------------------LysHisSerThrHisPheSerThrTrpValLeuTyrAsnCy 547

Qy  2068 CCTTCGCATTATACAAGTACCTTCTAAGCTTTGAAATTGGAAGCTTCTACAGTATGCA 2127
                  ||||||||
Db   547 sPheArgAlaIleMetLeuIlePheLeuMetSerGlyPheGluLeuGluLeuTyrAlaValHi 567

Qy  2128 CGAGTACTATTACATATATTAGTATCTCTCTTCGAATTCCTTTACGCATGGTGAGTCAAC 2187
                 ||||:::   :::   ||||||||||   :::|||||||
Db   567 sGluPheLeuTyrIleTyrTrpTyrProTyrGluPheGluLeuGlyPheGluValSerAl 587

Qy  2188 ATTGAGTCGTGTGCCCATGGCTCTCAAATGGCAGAGGAAAGGATAATGGAAGAGCAGAA 2247
                     :::
Db   587 aLeuThrArgThrGluAsnIleLeuLeuAlaGlnLeuTyrAlaGlnHisGlnSerGly 607

Qy  2248 AGGCCGTAGT------------------------AGTAAAACAAAAGAA 2274
                  ||||||
Db   607 sThrGlnSerGlyValGlySerGlyTyrAlaAlaLysAlaLaAlaLysProLysLysAs 627

Qy  2275 AAAAAAGAAAGTTCGCCCATTGAGCCGAGAGGAGGAGATCACAAATGAGCCAAGAGATATCAGAACAT 2334
                  |||                                     :::
Db   627 nLysLysThrGlnArgProTyrArgAlaGluIleValPheTyrHisAlaLeuLeuSerLe 647

Qy  2335 GTGTGCTGGAATGTTAAAACCATGGTAGCATTTGACATGGACGGGCAAAGTACGTAAAACC 2394
                  :::||||||
Db   647 uCysGlyGlyMetTyrLysAlaMetGlyAlaLeuThrLysAspGlyArgValArgLeuPr 667

Qy  2395 GAAGTTTGAGCTTGATAGTGGAACAAAGTTCGGTATGAACACAAGGTTTGCTTCCATTCAACAG 2454
                  :::|||
Db   667 oLeuSerLysPheAspAsnGluIleLeuArgTyrAsnArgArgPheLeuProPheAlaTh 687

Qy  2455 TCTGATGACCCCGCGGCTGCACTTACTTACAGTTCAAGGAAAATGTCTGAC---CTCAA 2511
                  |||
Db   687 rLeuThrSerProProProValSerTyrAlaGluPheLysAsnValArgGluHisMetMe 707

Qy  2512 TAAATATAGCCCTCCTCCTCGAACTGTATGGGCAGCTAGTGAAGCACTTTCA 2571
                  :::|||                       :::
Db   707 tArgHisSer------------------ValGluAspLeuTyrThrTyrAlaAlaLysHisPheAs 723

Qy  2572 ACAGGCAAAAATGATATTGGGAAAATATTCCTAACCCGGACCATGAGGTTAATAGAATTTT 2631
                  ||||
Db   723 pGlnAlaArgAsnValLeuGluSerIleGlnAsnProAspGlnGluMetLeuAspLeuIle 743

Qy  2632 AAAGGTTGCCAAACCCAAACTTTGTGTGTTATGAAGTTATTGGCAGGAGGACACAAAAAGGA 2691
                  |||
Db   743 uValLysArgGlnProGluPheAspPheSerLysHisSerTyrPheProIleLeuLysLe 763

Qy  2692 ATCTAAAGTTCCTCCTGAATTGATTTCGCTCATAAAATAATATTTCCGTTGTGGAAACT 2751
                  |||                              :::|||||||   :::||||
Db   763 uGlnIleLeuAlaArgThrAsnPheValValMetAsnValLeuAlaArgGlyHisGlnLysGl 783

Qy  2752 T 2752
         u
Db   783 u 783
```

SQ   Sequence   239 AA;

Alignment Scores:
Pred. No.:
Score: 7.99e-109
Score: 1237.00
Percent Similarity: 99.58%
Best Local Similarity: 99.58%
Query Match: 22.74%
DB: 23

Length: 239
Matches: 238
Conservative: 0
Mismatches: 0
Indels: 1
Gaps: 0

US-10-001-857-42 (1-3096) x ABG59816 (1-239)

```
Qy  475 GCGGGCGGCCGAGCCGGCGTCGTTATTTCCGTGGTCCGGACAGTGCGGTGGCGGGCGGGGTG 534
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AlaAlaAlaGluAlaAlaSerLeuPheProTrpSerGlyGlnCysValAlaAlaArgVal  20

Qy  535 ACCACGGGAGAAGGTAGGCGCATAATAGGTTATGAAAAGCTTCTCGTAGATGATGAAGGAGGA 594
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 ThrThrGlyGluValGlyIleMetValMetLysAlaSerValAspSerAspSerGly   40

Qy  595 TGGGGAGGTCAGTATGCCAGAAAAAATGAGGAAAAGCAATACAAACTGGGTGGACATTACC 654
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 TrpGluLeuSerMetProGluLysMetGluLysSerAsnThrAsnTrpValAspIleThr  60

Qy  655 CAAGATTTTGAAGAAGCTTGTGCTGAGAATTAAAGTTGGGAGGAACTACTTCATGATAAGCTA 714
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GlnAspPheGluGluAlaCysAlaGluAsn...LysLeuGlyLeuLeuHisAspLysLeu  80

Qy  715 TTTGGTCTTTTTGGAAGCCATGTGCTATTGAAAATGAGATCCCAAGATGGGATGCTGGC 774
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 PheGlyLeuPheGluAlaMetSerAlaIleGluMetMetAspProLysMetAspAlaGly 100

Qy  775 ATGATTGGAAAACCAAGTTAATCGAAAAGTTCTCAATTTGAACAAGCTATCAAGGATGGC 834
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  101 MetIleGlyAsnGlnValAsnArgLysValLeuAsnPheGluGlnAlaIleLysAspGly 120

Qy  835 ACTATTAAAATTAAAGATCTCACCTTGCCTGAACCTAGGAATTATGGATTCATAGTTTTT 894
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ThrIleLysIleLysAspLeuThrLeuProGluLeuArgAsnTyrGlyPheIleValPhe 140

Qy  895 TGCTGTTTGATAAACGTGGTTAGAAGGCCAATCACTGGCACAGACAGTATTACGTGCCTT 954
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  141 CysCysLeuIleThrTrpLeuGluGlyHisSerLeuAlaGlnThrValPheThrCysLeu 160

Qy  955 TACATTCATAATCCAGACTTATTAGAAAATCCTCGTATGAAAGGCTTTGCTCTGGGAATC 1014
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  161 TyrIleHisAsnProAspPheLeuGluAspProAlaMetLysGlyPheAlaLeuGlyIle 180

Qy 1015 TTGAAAATCTGGACAATTGCAAGGGAAAAAAGTAAATAAACTGCTCGTTTTTTGAAGAGGAA 1074
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 LeuLysSileCysAspSerIleAlaArgGluLysValAlaAsnLysLysAlaAlaValPheGluGlu 200

Qy 1075 GATTTTCAGTCAATGACTTATGGATTTAAAATGGCTAACAGTGTGACAGATCTTCGAGTT 1134
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  201 AspPheGlnSerMetThrTyrGlyPheLysMetAlaAsnSerValThrAspLeuArgVal 220

Qy 1135 ACAGGCATGCTAAAAAGATGTGAGGGGATGACATGCAAAGAAGAGAGTAAAGAGTACTCGA 1191
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  221 ThrGlyMetLeuLysArgCysGluGlyMetThrCysLysGluGluSerLysGluTyrSerArg 239
```

RESULT 5
AAU32414
ID   AAU32414 standard; Protein; 108 AA.
XX
AC   AAU32414;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Novel human secreted protein #2905.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune stimulation; anti-inflammatory; leukaemia.

XX
OS   Homo sapiens.
XX
FN   WO200179449-A2.
XX
PD   25-OCT-2001.
XX
DR   WPI; 2001-611725/70.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
PA   (HYSE-) HYSEQ INC.
XX
PR   18-APR-2000; 2000US-0552929.
PR   26-JAN-2001; 2001US-0770160.
XX
PT   Nucleic acids encoding a range of human polypeptides, useful in genetic
PT   vaccination, testing and therapy -
XX
PS   Claim 20; Page 608; 765pp; English.
XX
CC   The invention relates to novel human secreted polypeptides. The
CC   polypeptides and antibodies to the polypeptides are useful for
CC   determining the presence of or predisposition to a disease associated
CC   with altered levels of polypeptide. The polypeptides are also useful for
CC   identifying agents (agonists and antagonists) that bind to them. Cells
CC   expressing the proteins are useful for identifying a therapeutic agent
CC   for use in treatment of a pathology related to aberrant expression or
CC   physiological interactions of the polypeptide. Vectors comprising
CC   the nucleic acids encoding the polypeptides and cells genetically
CC   engineered to express them are also useful for producing the proteins.
CC   The proteins are also useful in genetic vaccination, testing and
CC   therapy, and can be used as nutritional supplements. They may be used to
CC   increase stem cell proliferation; to regulate haematopoiesis; and in
CC   bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC   immune suppression and/or stimulation; as anti-inflammatory agents; and
CC   in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC   sequences of novel human secreted proteins of the invention.
XX
SQ   Sequence   108 AA;

Alignment Scores:
Pred. No.:
Score: 2.28e-28
Score: 399.00
Percent Similarity: 96.39%
Best Local Similarity: 96.39%
Query Match: 7.13%
DB:

Length: 108
Matches: 80
Conservative: 0
Mismatches: 2
Indels: 2
Gaps: 0

US-10-001-857-42 (1-3096) x AAU32414 (1-108)

```
Qy 2944 ACCACTTTAAAACAGCAGTTTTCATTCATGTCATATAAGGCTGTCAAACTAAAAGTATTC 2885
        ||||||||||||||||||||||||||||||||||||||| |||||||||| ||| |||
Db    1 ThrThrLeuLysGlnGlnPheSerPheMetSerTyrLysAlaValLysLeuLysValPhe  20

Qy 2884 CTTATAATGAGTTGTTACCCAAGAAAATCCATTCTCACTTCCCGTGATGTGGAGCCTGGTG- 2826
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 LeuIleMetSerCysTyrProArgAsnProSerHisPhePro***CysGlyAlaTrpVal  40

Qy 2825 ATGTGCCCTCTAAGAGTTGGGTCGAAAGAAGGACTTGCCCCTTTATGGCCACCTCCCCA 2766
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 MetCysProLeuArgValGlySerGluArgArgLeuCysProPheMetAlaThrSer-Gl  60

Qy 2765 GTCTCTCTCCAAACAAGTTTCACACACAGGAAAAATCAAATATTTATGAGCAGAGAAATCAAATTCAG 2706
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 nSerLeuSerAsnLysPheHisAsnArgLysIlePheMetSerArgGluIleLysPheAr  80

Qy 2705 GAGGAAC 2699
        |||||||
Db   80 gArgAsn  82
```

RESULT 6
ABB93650

ID   ABB93650 standard; Protein; 681 AA.
XX
AC   ABB93650;
XX
DT   31-MAY-2002 (first entry)
XX
DE   Herbicidally active polypeptide SEQ ID NO 2861.
XX
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO20210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP09892.
XX
PR   28-AUG-2001; 2001WO-EP09892.
XX
PA   (FARB) BAYER AG.
XX
PI   Tietjen K, Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   organisms -
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   Identifying plant target proteins for herbicidally active compounds,
XX
PS   Claim 5; SEQ ID NO 2861; 261pp + Sequence Listing; English.
XX
CC   useful as herbicides.
CC   useful for identifying modulators. The identified modulators are
CC   are useful for identifying them
CC   sequences or nucleic acids encoding them
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   greater by a factor of 3 than the E-value of most similar non-plant
CC   suitable search parameters, where plant sequences having an E-value
CC   with nucleic acid or amino acid sequences from non-plant organisms using
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   The invention relates to identifying target proteins

SQ   Sequence   681 AA;

Alignment Scores:
Pred. No.:
Score: 158.50   Matches: 52
Percent Similarity: 38.65%   Conservative: 11
Best Local Similarity: 31.90%   Mismatches: 60
Query Match: 2.88%   Indels: 40
DB: Gaps: 6

US-10-001-857-42 (1-3096) x ABB93650 (1-681)

Qy   608 ATACTAGAGCTCCCATCCTGAATCTTCATCTACTACAGAAGCTTTCATAAACCATTATGCCT 549
            ||||:|||||||||||:::|||| ||:||||||||||||||||||||||||||
Db   60 ValValSerProProSerSerSerSerPro----------------------Pro 72

Qy   548 ACTTCTCCGTGGTCACCGCGCGCGCGCCGGACCACTGTCCGG------------------ 509
            :::|||||||| |::||||||||||||||||||||
Db   73 SerProProValIleThr-SerProProThrValAlaSerSerProProProVa 92

Qy   508 ------------------ACCACGGAAATACGACGCCCTGGCGCGCG----- 476
                             :::||||||||||||||||||||
Db   92 1ValIleAlaSerProProSerThrProAlaThrProProAlaProProGlnTh 112

Qy   475 -----------CCGCCCCTCCCCCTCCAGCCCGACCGGCAGCCCGCTGCACGCANTGCGTAT 423
            ||||||||||||||||||||||||
Db   112 rValSerProProProProProAspAlaSerProSerProAlaProThrThrThrAs 132

Qy   422 GCCCGGGACGGCGGGACCCGCCCCCGCCGCCCTCCTCCCTTACTCCCGGGGGGGACTTGCGCTCCCCGGC 363
            ||| |||| |:: ||||||||| |||||||| ::: |||
Db   132 nPro----------ProProLysProSerProSerProSerProProGlyGluThrProSerProPr 149

Qy   362 CACACCCCTCCTCCGCCACCGCCTCCGTTCGGCCGATGGCGGCGCCGCCTTCTCTTCTCGG 303
            |||||||||||||||||| ||||||||||||
Db   149 oGlyGluThrProSerProProLysPro----------SerProSerThrProTh 164

Qy   302 GAGAGTTAGGAGACGATCGGGAGACCGGAGACCGGGAAGTTATCTTTTTTTCACCCTCTGTCCTG 243
            |||||||||| ::: |||||
Db   164 rProThrThrThrThrSerProPro-------------ProProAlaThrSe 178

Qy   242 GGTAGTGGGGCGGTTCCCGGGAAAAGAGGCGGAGCCCGGAGTCTCAGAGCCCCGTCTCGC 183
            ||||| ||| |||:|||
Db   178 rAlaSerProSerSerAsnProThrAspProSerThrLeuAlaProProThrPr 198

Qy   182 GCTGCCC 176
            |||:|||
Db   198 oLeuPro 200

RESULT 7
ABG14149

ID   ABG14149 standard; Protein; 156 AA.
XX
AC   ABG14149;
XX
DT   18-FEB-2002 (first entry)
XX
DE   Novel human diagnostic protein #14140.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO20017S067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-G393362/73.
DR   N-PSDB; AAS78336.
XX
PT   biodiversity -
PT   responsible for genetic disorders or other traits and to assess
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   New isolated polynucleotide and encoded polypeptides, useful in
XX
PS   Claim 20; SEQ ID No 44508; 103pp; English.
XX
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   The polypeptide and polynucleotide sequences have applications in
CC   disorders involving aberrant protein expression or biological activity.
CC   imaging of sites expressing (II), and (II) are useful for treating
CC   a food supplement. (II) and its binding partners are useful in medical
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   to restore normal activity of (II) or to treat disease states involving
CC   for identifying expressed genes. (II) is useful in gene therapy techniques
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   and gene mapping, and in recombinant production of (II). The
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   The invention relates to isolated polynucleotide (I) and

```
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
CC
SQ   Sequence   156 AA;
XX

Alignment Scores:
Pred. No.:
Score:                 1.61e-05
Percent Similarity:    34.66%
Best Local Similarity: 29.55%
Query Match:           2.79%
DB:

US-10-001-857-42 (1-3096) x ABG14149 (1-156)

Qy     556 TTATCCTTACTTCTTCCCGTGGTCACCCGGCCCACTGTCCGGACCACGGAAATA 497
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 LeuCysSerLeuSerSerLeuProValProProProProPro------------ 12

Qy     496 ACGAGCCGCCTCGGGCCCCGTCCTCTCAGCCCACCCGACCGGCAGCGT 437
           ||||||||||||||||||||||||||||||||||||||||||
Db      13 --ProProProPheProProProProPro-------------------- 22

Qy     436 GCACGGCATGCGTATGCCCGGGGACGGGCGGCCCCCCCCCTCTCCTTACTCCCGGGGGACT 377
           |||:|||||||||||||||::|||::|||:||||||||
Db      23 --ProProProProPheProProProSerProValProLeuProPro----- 37

Qy     325 TGGGGCGGCGCCTTCTTCTTCTCGGAGATTAGAGACGATCGCGACGGGAAGTATTATCTTT 266
           |||||||||
Db      38 --SerProArgSerProProValSerProProProHisSerPheGlnGlyArg 55

Qy     376 TCGCGGCTCCCCGGGCCACACCCCCTCCTTCGCCACGCGCTTCCGTTC---- GGCCGA 326
           ||||
Db      56 SerProProSerGlyLeuGlySerSerAlaGluProTrpLeuArgProGlyThrTrpVal 75

Qy     265 TTTCACCCT---------------CTGTCGGAGTCCTGGGTAGTGGGCGGTTCCCGAAAGAGGCG 215
           |||                                             |||
Db      76 ***ProProProLeuThrPheSerGlnGlnGluAlaSerSerAlaGlnLys***Asn 95

Qy     214 GAGCGCCGGAGTCTCA---GAGCCGCGCCCCCGTCTGCGCTCCTGGGAATTCCTTAGAA 158
           |||
Db      96 CysProGlyGluSerProGlnProValProValThrValPro------------------- 109

Qy     157 AGGTGACGCGAAGGGGGAGGAAGGAAACCGTGGAGTTTGCGTAGACCTTAGTGGTTGCGAGC 98
           |||
Db     109 ------------------------------------------------------- 109

Qy      97 ACGAACGAGCGAGCCAAGCTCCGAGAGCAAACAGCGTGCTACTGCTGCCT 50
           |||||||||||||||||||||||||||||||
Db     109 ------------------------------------------ 109

Qy     110 --GluArgGlnProAspAlaAlaProAsnHisSerLeuLeuLeuPro 124
           |||||||||||||
Db
```

```
RESULT 8
ID   ABB91504 standard; Protein; 708 AA.
XX
AC   ABB91504;
XX
DT   31-MAY-2002   (first entry)
XX
DE   Herbicidally active polypeptide SEQ ID NO 715.
KW   Herbicidal; plant; agriculture; herbicide.
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
```

```
PF   28-AUG-2001; 2001WO-EP09892.
XX
PR   28-AUG-2001; 2001WO-EP09892.
XX
PA   (FARB ) BAYER AG.
XX
PI   Tietjen K, Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PS   Claim 5; SEQ ID NO 715; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   with nucleic acid or amino acid sequences from non-plant organisms using
CC   suitable search parameters, where plant sequences having an E-value
CC   greater by a factor of 3 than the E-value of most similar non-plant
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   are useful for identifying modulators. The identified modulators are
CC   useful as herbicides.
XX
SQ   Sequence   708 AA;
XX

Alignment Scores:
Pred. No.:
Score:                 4.11e-05
Percent Similarity:    39.38%
Best Local Similarity: 30.57%
Query Match:           2.78%
DB:

US-10-001-857-42 (1-3096) x ABB91504 (1-708)

Qy     608 ATACTGAGCTCCCATCCTGAATCGTCATCATCTACAGAAGCTTTCATAAACCATTATGCCT 549
           ::::::                                       :::||||
Db     101 ValIleProSerProProSerProAlaSerProProProAlaLeuValProProLeuPro 120

Qy     548 ACTTCTCCC--------------------------GTGGTCACCCGGCCGCCACG 519
           :::||||||||                      |||||||||||
Db     121 SerSerProProProAlaSerValProProAlaProArgPro--SerProSerProProI11 140

Qy     518 CACTGTCCGGACC--------------ACGGAAATAACGACGCCGGCCTCGGCGCCG-- 476
           |||||||||||                 :::|
Db     140 eLeuValArgSerProValArgProIleGlnSerProProProProSe 160

Qy     475 -----------------CCGCCCCCCTCCCTCTTCAGCCCAGCCGCCGACCGGCCGTGC 435
           ||                                         :::||||
Db     160 rAspArgProThrGlnSerProSerProProSerProProSerProSerGluArgPr 180

Qy     434 ACGCATG----------------CGTATGCCCGGGACGGCGCGACCCGCCCTC 396
           |||                                |||
Db     180 oThrGlnSerProSerProSerGluArgProThrGlnSerProSerProProSe 200

Qy     395 TCCCTTACTCCCGGGGACTTCGGCGTCCCCGGCCACACCCCCTCCCTGCCACGCGCTCC 336
           |||          :::        ||||||||
Db     200 rProProSerProSerAspArgProSerGlnSerProPro------ProProGl 218

Qy     335 GTTCGGCCGATGGGCGCGGCCCGCCTTCTTCTTCTCGGAGATTAGAGACGATCGCGACCGGA 276
           :::
Db     218 uAspThrLysProGlnProProArg-------------ArgSerProAsnSerPropr 233

Qy     275 AGTTATCTTTTTCACCCTCTGTCGGAGTCCTGGGTAGTGGGCGGTTCCCGAAAGAGGC 216
           |||||||||||
Db     233 oProThrPheSerSerPro-ProArgSerProProGluIleLeuValProGlySerAsnA 253

Qy     215 GGAGCCCGGAGTCTCAGAGCCCGCGCCCCCGTCTGCGCTCCTGGGAATTTCTTAGAAAG 156
           ||||||                             |||||||
Db
```

```
Db
Qy  253  sn-------ProSerGlnAsnAsnProThrLeuArgProProLeu---------------- 265
Qy  155  GTGACGGCGAAGGCGGGAAGGAAGGAAGAAACCGTGGAGGTTTG 121
         |||  ||||:::
Db  266  --AspAlaProAsnSerThrAsnAsnSerGlyIle 276

RESULT 9
ABUS2794
ID   ABUS2794 standard; Protein; 120 AA.
XX
AC   ABUS2794;
XX
DT   14-APR-2003  (first entry)
XX
DE   Human transmembrane protein DKFZphfbr2_82g14 homologue #2.
XX
KW   Human; gene therapy; vaccine; disease treatment; detection.
XX
OS   Homo sapiens.
XX
PN   WO200112659-A2.
XX
PD   22-FEB-2001.
XX
PF   18-AUG-2000; 2000WO-IB01496.
XX
PR   18-AUG-1999; 99US-0149499.
PR   28-SEP-1999; 99US-0156503.
XX
PA   (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX
PI   Wiemann S;
XX
DR   WPI; 2001-327840/34.
XX
PT   Nucleic acids having the sequences of clones isolated from libraries of
PT   different human tissues, useful in recombinant DNA methodologies -
XX
PS   Example III; Page 359; 1095pp; English.
XX
CC   This invention describes novel polynucleotides and polypeptides isolated
CC   from human cDNA libraries which can be used for gene therapy or in
CC   vaccines. The polynucleotides of the invention and antibodies encoded by
CC   them may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate polypeptide expression. The products of the
CC   invention may also be used to identify modulators of expression and
CC   activity and to down regulate expression and activity. The antibodies of
CC   the invention may also be used as diagnostic agents for detecting the
CC   presence of polypeptides in samples. This sequence represents a homologue
CC   of a polypeptide described in the disclosure of the invention.
XX
SQ   Sequence  120 AA;

US-10-001-857-42 (1-3096) x ABUS2794 (1-120)

Alignment Scores:
Pred. No.:

Score:                 6.5e-05          Length:       120
Percent Similarity:    146.50           Matches:      44
Best Local Similarity: 38.93%           Conservative: 7
Query Match:           33.59%           Mismatches:   53
DB:                    2.67%            Indels:       27
                       22               Gaps:         5

Qy  541  CCGTGGTCACCGGCGCGCGCGCCCCACGCACTGTCCGGACCAAATAACGACGCCCGCCT--- 485
         ||||||| ||  ||:||||||||||||||||||||||||||::
Db  6    ProProGlyProProGlyGlnValLeuProProLeuAlaGlyProProAsn 25
Qy  484  ---------CGGCCGCGCCGCCGCCCCTCCCCCCTCTCAGCCCGACCGGCAGGTGCA 434
         :::||||||||||:||
Db  26   ArgGlyAspArgProProProValLeuPheProGlyGlnProPheGlyGlnProPro 45
Qy  433  CGCAATGCGTATGCCCGGACAGCGGGACCCCGCCCCTCCTTACTCCCGGGGACTTCG 374
```

```
Db
Qy  46   LeuGlyProLeuPro---------------ProGlyProProProProProValProGlyProGlyTyrGly 61
Qy  373  CGCTCCCCCGGCACCACCACCCCCTCCC------TCGCCACCGCCTCCGTTGGCGATGG--- 323
         ||||     ||    ::
Db  62   ProProProGlyProProProGlnGlnGlyProProGlnGlyProProProGlyProPhePro 81
Qy  322  ---CGGCGCGCCTTCTCTTCGCGAGATTAGAACGATCGCGAGACCGGAAGTTATCTTT 266
         ||||||| :::
Db  82   ProArgProProGlyProLeuGlyProProLeuThrLeuAlaProProHisLeuPro 101
Qy  265  TTTCACCCTCTGTCGGAGTCCTGGGTTAGTGGCGGTTCCCGGAAAGAGGCGGAGCCCGGA 206
         ||:|||| ||||||||||
Db  102  GlyProPro--------- 106
Qy  205  GTCTCAGAGCCCGCCCCCCGTCGCGCTGCCCTCT 173
         ||||||||||| |||:||
Db  107  AlaProProAlaProHisValAsnProAla 117

RESULT 10
AAM52322
ID   AAM52322 standard; Protein; 574 AA.
XX
AC   AAM52322;
XX
DT   18-JAN-2002  (first entry)
XX
DE   WASP homolog protein.
XX
KW   Actin polymerisation; Ena/VASP; vasodilator-stimulated phosphoprotein;
KW   metastatic cancer; parasitic infection; cytotoxic; WASP.
XX
OS   Schizosaccharomyces pombe.
XX
PN   WO200171356-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-FR00843.
XX
PR   22-MAR-2000; 2000FR-0003637.
XX
PA   (CNRS ) CENT NAT RECH SCI.
PA   (CURI-) INST CURIE.
XX
PI   Fradelizi J; Friederich E, Golsteyn RM, Louvard J, Noireaux V;
PI   Sykes C;
XX
DR   WPI; 2001-639148/73.
XX
PT   Identifying modulators of actin polymerization, potentially useful for
PT   treating tumor metastasis and parasitic infection, using proteins that
PT   contain Ena/VASp binding sites -
XX
PS   Claim 13; Pages 107-109; 109pp; French.
XX
CC   The present invention relates to a method for identifying modulators of
CC   actin polymerisation. The method involves using proteins that contain at
CC   least one binding motif for proteins of the Ena/VASP
CC   (vasodilator-stimulated phosphoprotein) family in the preparation of
CC   reagents for identification/screening of molecules that modulate
CC   formation of the actin cytoskeleton. The proteins used in the method
CC   (i.e. the proteins with binding motif(s) for Ena/VASP proteins) do not
CC   bind to the Arp2/3 protein complex. The modulators identified by the
CC   method are potentially useful for treating disorders of actin
CC   polymerisation, e.g. metastatic cancer or parasitic infection; and as
CC   cytotoxic agents. The present sequence one such protein with binding
CC   motif(s) for Ena/VASP proteins, which was used in the method of the
CC   present invention.
XX
SQ   Sequence  574 AA;

Alignment Scores:
Pred. No.:             0.000153         Length:       574
```

Score: 146.50   Matches: 53
Percent Similarity: 39.41%   Conservative: 14
Best Local Similarity: 31.18%   Mismatches: 64
Query Match: 2.67%   Indels: 39
DB: 22   Gaps: 8

US-10-001-857-42 (1-3096) x AAM52322 (1-574)

```
Qy   544 CTCCCGTGGTCACCCGCCGCCGCCACGCACTGTCCGGACCACGGAAATAACGACGCCGCCT 485
             |||| |||||||||  ||||     |||     ||||||    ||||||||| |||||
Db   336 LeuProProProProProProAlaProArgSerAsnAlaAlaGlySerIleProLeuProPro 355

Qy   484 CGG-------------CCGCGCGCCGCCCCCTTCCCAGCCCACCCGACCGGCAGCGT 437
             :::           ::||||||||||||||||||||||||||||||||| |||||
Db   356 GlnGlyArgSerAlaProProProProProProArgSerAla---ProSerThrGlyArg 374

Qy   436 GCACGCATGCGTATGCCCGGGACCGC-------------GACCCCGCGCCCTCTCCTTA 389
             |||||||||||||||||||||||||             ||||||||||||||||| ||
Db   375 GlnProProProLeuSerSerSerArgAlaValSerAsnProProAlaProProProAla 394

Qy   388 CTCCCGGGGACTTCGCCGCTCCCCGGGCCACACCC------------------------- 356
             |||||||||||| ||||||||||||||||||||||
Db   395 IleProGlyArgSer---AlaProAlaLeuProProLeuGlyAsnAlaSerArgThrSer 413

Qy   355 ------------------------CCTCCCTGCCACCGCCTGCCTTCGTTCGGCCGA 326
                                     |||||||||||||||||||||||||||||||||
Db   414 ThrProProValProThrProHisProSerLeuProProSerLeuProProSerLeuProPro 433

Qy   325 TGGCGCGCCGCTTCTCTTCGCGGAGATTAGACAGATGCGAGACGCCGGAAGTTATCTTT 266
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   434 SerAlaAlaProProSerLeu------ProMetGlyAlaProAlaProAlaAlaProAlaAla 451

Qy   265 TTTCACCCTCTCGTCGGAGTCCTTGGGTAGTGGGC 227
             |||||||||||||||||||||||||||||||||||
Db   472 AlaAlaAlaProAlaAlaProAlaProAlaProAlaAla---ProValAlaSer 490

Qy   166 TTCTTAGAAAGTGACGCGAAGGCGGAAGG 137
             |||||||||||||||||||||||||||||
Db   491 IleAlaGlyLeuProGlnGlnAspGlyArg 500
```

Amino acid sequence of a yeast WASP protein homologue.

RESULT 11

AAG67370
ID   AAG67370 standard; Protein; 574 AA.
XX
AC   AAG67370;
XX
DT   13-NOV-2001 (first entry)
XX
DE   Amino acid sequence of a yeast WASP protein homologue.
XX
KW   Wiskott-Aldrich syndrome protein; WASP; actin cytoskeleton;
KW   cell motility; actin polymerisation; cancer; parasite infection;
KW   embryonic development; immune response; wound repair.
XX
OS   Schizosaccharomyces pombe.
XX
XX
PN   WO200144292-A2.
XX
PD   21-JUN-2001.
XX
PF   15-DEC-2000; 2000WO-FR03569.
XX
PR   16-DEC-1999; 99FR-0015900.
XX
PA   (CURI-) INST CURIE.
PA   (CNRS) CENT NAT RECH SCI.
XX
PI   Noireaux V, Prost J, Sykes C, Friederich E, Golsteyn RM;
PI   Louvard D;
XX
PT   New fragments of WASP family proteins, useful for detecting and
PT   identifying modulators of actin cytoskeleton formation, potential
PT   anticancer and antiparasitic agents -
XX
PS   Claim 14; Fig 8; 162pp; French.
XX
CC   The present sequence represents a WASP (Wiskott-Aldrich syndrome
CC   protein) homologue. Peptide fragments of WASP-family proteins of
CC   eukaryotic cells are used to prepare reagents for detecting compounds
CC   that inhibit or stimulate formation of the actin cytoskeleton, and
CC   thus inhibit or stimulate cell motility. The peptides are used to
CC   detect and identify compounds which are potentially useful for treating
CC   diseases associated with dysfunction of actin polymerisation,
CC   particularly metastatic cancer and parasite infection; as cytotoxic
CC   agents for inhibiting/stimulating formation of the actin cytoskeleton
CC   and for detecting side-effects, on actin polymerisation, of
CC   pharmaceuticals. By modulating actin polymerisation, these compounds
CC   affect cell motility, embryonic development, the immune response and
CC   wound repair.
XX
DR   WPI; 2001-536241/59.
DR   N-PSDB; AAH77922.
XX
SQ   Sequence   574 AA.;

Pred. No.:   0.000153
Alignment Scores:
Score: 146.50   Length: 574
Percent Similarity: 39.41%   Matches: 53
Best Local Similarity: 31.18%   Conservative: 14
Query Match: 2.67%   Mismatches: 64
DB: 22   Indels: 39
                                     Gaps: 8

US-10-001-857-42 (1-3096) x AAG67370 (1-574)

```
Qy   544 CTCCCGTGGTCACCCGCCGCCGCCACGCACTGTCCGGACCACGGAAATAACGACGCCGCCT 485
             |||| |||||||||  ||||     |||     ||||||    ||||||||| |||||
Db   336 LeuProProProProProProAlaProArgSerAsnAlaAlaGlySerIleProLeuProPro 355

Qy   484 CGG-------------CCGCGCGCCGCCCCCTTCCCAGCCCACCCGACCGGCAGCGT 437
             :::           ::||||||||||||||||||||||||||||||||| |||||
Db   356 GlnGlyArgSerAla---ProAlaLeuProProLeuGlyAsnAlaSerArgThrSer 413

Qy   388 CTCCCGGGGACTTCGCCGCTCCCCGGGCCACACCC------------------------- 356
             |||||||||||| ||||||||||||||||||||||
Db   375 GlnProProProLeuSerSerSerArgAlaValSerAsnProProAlaProProProAla 394

Qy   436 GCACGCATGCGTATGCCCGGGACCGC-------------GACCCCGCGCCCTCTCCTTA 389
             |||||||||||||||||||||||||             ||||||||||||||||| ||
Db   356 GlnGlyArgSerAlaProProProProProProArgSerAla---ProSerThrGlyArg 374

Qy   355 ------------------------CCTCCCTGCCACCGCCTGCCTTCGTTCGGCCGA 326
                                     |||||||||||||||||||||||||||||||||
Db   414 ThrProProValProThrProHisProSerLeuProProSerLeuProProSerLeuProPro 433

Qy   325 TGGCGCGCCGCTTCTCTTCGCGGAGATTAGACAGATGCGAGACGCCGGAAGTTATCTTT 266
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   434 SerAlaAlaProProSerLeu------ProMetGlyAlaProAlaProAlaAlaProAla 451

Qy   265 TTTCACCCTCTCGTCGGAGTCCTTGGGTAGTGGGC 227
             |||||||||||||||||||||||||||||||||||
Db   452 SerAlaProSerLeu------ProMetGlyAlaProAlaProAlaAlaProAla 471

Qy   226 CGGAAAGAGGCGGAGCCCGGAGTCTCAGAGCCCCGGGCCCGCCCGTCTGGGAAT 167
             |||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   472 AlaAlaAlaProAlaAlaProAlaProAlaProAlaAla---ProValAlaSer 490

Qy   166 TTCTTAGAAAGTGACGCGAAGGCGGAAGG 137
             |||||||||||||||||||||||||||||
Db   491 IleAlaGlyLeuProGlnGlnAspGlyArg 500
```

RESULT 12

ID   ABG79529
AC   ABG79529;
XX
DT   15-NOV-2002 (first entry)
XX
DE   Human tumour suppressor protein #15.
XX
KW   Human; cancer; tumour suppressor protein; breast cancer; lymphoma;
KW   adenocarcinoma; leukaemia; melanoma; myeloma; sarcoma; teratocarcinoma;
KW   prostate; cervix; liver; ovary; adrenal gland; heart; brain; lung; colon;
KW   placenta; skeletal muscle; synovial membrane; tonsil; kidney; uterus;
KW   skin; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200264775-A1.
XX
PD   22-AUG-2002.
XX
PF   12-FEB-2002; 2002WO-AU00137.
XX
PR   12-FEB-2001; 2001AU-0003052.
PR   12-FEB-2001; 2001AU-0003053.
PR   12-FEB-2001; 2001AU-0003054.
XX
PA   (BION-) BIONOMICS LTD.
XX
PI   Callen DF, Powell JA, Kremmidiotis G, Gardner AE, Whitmore SA;
XX
DR   WPI; 2002-657597//70.
DR   N-PSDB; ABS64842.
XX
PT   New tumour suppressor genes identified at 16q24.3, useful for
PT   identifying or obtaining full-length human genes involved in the
PT   tumourigenic process, or in diagnosing or treating cancer (e.g. breast
PT   or cervix cancer) through gene therapy -
XX
PS   Claim 15; Page 138-139; 218pp; English.
XX
CC   The invention relates to a tumour suppressor gene and the protein it
CC   encodes. The genes are useful for identifying tumourigenic and/or
CC   full-length human genes involved in the tumourigenic process. The genes
CC   the polypeptides and antibodies to the polypeptides are useful in
CC   diagnosing cancer, in establishing the prognosis of a patient diagnosed
CC   with cancer, or in treating cancer through gene therapy. In particular,
CC   the cancer is breast cancer. The drugs that restore tumour suppressor
CC   activity are useful for treating breast cancer or for manufacturing a
CC   medicament for the treatment of breast cancer. The cancer may also be
CC   adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma,
CC   teratocarcinoma, or cancer of the prostate, cervix, liver, ovary, adrenal
CC   gland, heart, brain, lung, placenta, skeletal muscle, synovial membrane,
CC   tonsil, kidney, colon, uterus or skin. This sequence represents a human
CC   tumour suppressor protein of the invention.
XX
SQ   Sequence 416 AA;

US-10-001-857-42 (1-3096) x ABG79529 (1-416)

Alignment Scores:
Pred. No.:
Score:

0.000143	Length:	416
146.00	Matches:	62
Percent Similarity: 37.62%	Conservative:	17
Best Local Similarity: 29.52%	Mismatches:	62
Query Match: 2.66%	Indels:	69
	Gaps:	11

Qy   541 CCGTGGTCACCCGCGCCCGCCACGCACTGTCCGACC------------------------ 506
         ||| ||| |||||||||||||| ::||
Db   127 ProArgProProGlyProSerProGlnSerAspProGlnSerArgArgGlyPro 146

Qy   505 --ACGGAAATAACG 494

Db   147 ProAlaThrGlnAlaAsnProGluProProSerAlaProTyrGlyArgProLysLeuArg 166
XX

Qy   436 GCACGCCATGGTATGCCCGGGACGCGGGAC--------CCCGCCCCTCCTTACTCCG 383
         |||||||| | |||| ||||| :||||| ||||||:|||| |::::
Db   167 ArgProProArgProProGluProMetGlyProSerArgProProProArgAlaPro 186

Qy   382 GGGACTTGCGCCTCCCCGGCCACACCC-------------CCTCCGTGCCACCGCCTCCG 335
         |||:|| ||| ||| || |||| ||| |||| || || |||||||
Db   187 LysAspProArgLeuProArgProArgArgProAlaProAlaProArgProProIle-- 205

Qy   334 TTCGGCCGATGGCGGCGGCGCCT--------------------TCTCTTCGGAGATTAGAG 293
         ||||| |||||| | || ||::||||| :|
Db   206 GlyProSerArgSerProArgAlaProLysAspProArgProProGlyProAspProPro 225

Qy   292 ACGATCGCGCGGAGACCGGAAGTTATCTTTTTTCAC----------------------- 260
         |||||||||||| ||| ||| :||| :||
Db   226 ------------GlyProGluGlyProArgHisSerSerValAlaGlyLysGlyLeu 241

Qy   259 ---CCTTCGTCTGGGAGTCCTGGGTAGTGGGGTTCCGGAAAAGAGGAGCGGAGACCCGGA--- 206
         ||| ||||| || ||| |:::|
Db   242 GlnProAlaAsnProArgGluCysProValHisAlaProThrThrAlaProArgProPro 261

Qy   205 --------------------------------------GTCTCAGAGCCCGGCC------- 182
         | ||||:||||:||
Db   262 AlaProLeuThrValLeuProGlySerAlaThrAlaGlnArgArgGlyAla 281

Db   282 GlyAlaAlaHisAlaAlaArgGlyArgValSerArgProAlaArgProLysProGlnAla 301

Qy   181 CTGCCCTCTGGGAATTTCTTAGAAAGGTGACGGGAAAGGGGGAAGGAAAACCGTGGAGTTT 122
         ::||| ||| ||| || |:||||| |||:|| |||||||| ||||||||||| ||||||
Db   302 GlyGlyAlaGlyGly-AlaGlySerGlyProAlaLysLeuArgLeuArgGlyThrAlaGlyGly1 321

Qy   121 GCGTAGACCTTTAGTGTTGCGAGCACGA 94
         ||| ||||| ||
Db   321 uProAlaPro---------ArgAlaArg 327

XX
CC   The invention relates to identifying target proteins
CC   (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC   aligning and comparing nucleic acid or amino acid sequences from plant
CC   with nucleic acid or amino acid sequences from non-plant organisms using
CC   suitable search parameters, where plant sequences having an E-value
CC   greater by a factor of 3 than the E-value of most similar non-plant
CC   sequences are selected. The polypeptides or nucleic acids encoding them
CC   are useful for identifying modulators. The identified modulators are
CC   useful as herbicides.
XX
SQ   Sequence   694 AA;

US-10-001-857-42 (1-3096) x ABB92421 (1-694)

Alignment Scores:

Score:	0.000211
	145.50
Percent Similarity:	41.48%
Best Local Similarity:	32.59%
Query Match:	2.65%
DB:	23

Length:	694
Matches:	44
Conservative:	12
Mismatches:	46
Indels:	33
Gaps:	6

```
Qy  541 CCGTGGTCACCCGGGCCCCGCCACGCACTGTCCGGACCAGGAAATAACGACGCG------ 488
            |||                  :::||||| ||||||||
Db  104 ProLeuThrProSerProProAlaIleThrProSerProLeuThrProSerPro      123

Qy  487 ------CCTCGGCGCGCGGCCGCGCCCCCTCCCTCTCCAGCCCAGCCC 449
              |||||||||||
Db  124 LeuProProSerProThrProSerProProSerProSerIleProSerPro 143

Qy  448 GACCGGCAGCGTGCACGCATGCGTATGCCCGGGACGGCGACCCGCGCCCCC---TCTCCC 392
            ||                                          :: ||||||
Db  144 ------ProLeuThrProSerProProSerProProSerPro            154

Qy  391 TTACTCCCGGGG------ACTTCGCGCTCCCGCCCACACCCCCTCCCTCG 347
            |||                   ::|||||:||:||||||||
Db  155 LeuArgProSerSerProProProSerProSerProAlaThrProSer 174

Qy  346 CCACCGCGCTCCGTTCGGCCGATGGCCGGCCGCCTTCTCTTCTCCGGAGATTAGAGACGATC 287
            |||||||:||||||||                        :::
Db  175 ProProProProSerProThrProSerProProPro--------ArgValGlySerLeu 190

Qy  286 GGCGAGACCGGAAGTTATCTTTTTCACCCTCTGTCTGGGGTAGTGGGCGGGTTCC 227
            |||                                    :::
Db  191 SerProPro----------ProAlaSerProSerGlyGlyArgSerProSer 205

Qy  226 CGGAAAGAGGCGGAGCCCCGAGTCTCAGAGCCCGCCCCCGTCTCGCG 182
            ||||||||||
Db  206 ThrProSerThrThrProGlySerSerProProAlaGlnSerSer 220

Qy  206 ThrProSerThrThrProGlySerSerProProAlaGlnSerSer 220
```

RESULT 14
AA005346
ID   AA005346 standard; Protein; 149 AA.
XX
AC   AA005346;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polypeptide SEQ ID NO 19238.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO20016483S-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI85277.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 19238; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AA193841) and
CC   the encoded proteins (AA000010-AA013910) that exhibit activity relating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub//published_pct_sequences.
XX
SQ   Sequence   149 AA;

U-10-001-857-42 (1-3096) x AA005346 (1-149)

Alignment Scores:

Score:	0.000102
	145.00
Percent Similarity:	47.06%
Best Local Similarity:	42.35%
Query Match:	2.64%
DB:	

Length:	149
Matches:	36
Conservative:	4
Mismatches:	33
Indels:	12
Gaps:	3

```
Qy  553 TGCCTACTTCTCCCGGTGGTCACCCGCGCGCCACTGTCCGGACCACGGAAATAACG 494
            |||||:||:|:  |||    ||        |||:||:||||||  :::||||||
Db   74 CysLeuSerLeuPro--------------ProHisThrGlySerSerSerIleThr 89

Qy  493 ACGGCCG---------CCTCGGCGCGCGGCCGCGCCCCCTCTCAGCCCAGCGACCGG 443
            ||:||||         ||||||||||||||||||||||||||||||||||||||
Db   90 LeuProProLeuLeuProHisArgProHisProProProHisSerHisProMetAla 109

Qy  442 CAGCGTGCAGCGCATGGGTATGCCCGGGACGGGGACCCGGCGCCCCTCCTTACTCCG 383
            :::||||  ||  |||||||||||||||||||||||||||||  ||||||:||
Db  110 ThrLeuSerArgProProTyrProProProHisProProHisProLeuProPro 129

Qy  382 GGGACTTCGCGCTCCCGGCCACACCCCGCCCTCGCCACGGCGCTCCGTTCGGCCGATGG 323
            |||            |||||||||||||||||||||||||||||||||||||||||||
Db  130 SerLeuSerProProAlaLeuSerProProProProProProProPro--------- 145

Qy  322 CGGGCCGCCTTCTCTT 308
            ||| |||  | |
Db  146 ---ProProSerLeu 149
```

RESULT 15
AAM52317
ID   AAM52317 standard; Protein; 559 AA.
XX
AC   AAM52317;
XX
DT   18-JAN-2002 (first entry)
XX
DE   Human Scarl protein.

```
XX
KW   Actin polymerisation; Ena/VASP; vasodilator-stimulated phosphoprotein;
KW   metastatic cancer; parasitic infection; cytotoxic; Human; Scarl.
XX
OS   Homo sapiens.
XX
PN   WO20017135G-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-FR00843.
XX
PR   22-MAR-2000; 2000FR-0003637.
XX
PA   (CNRS ) CENT NAT RECH SCI.
PA   (CURI-) INST CURIE.
XX
PI   Fradelizi J, Friederich E, Golsteyn RM, Louvard D, Noireaux V;
PI   Sykes C;
XX
DR   WPI; 2001-639148/73.
XX
PT   Identifying modulators of actin polymerization, potentially useful for
PT   treating tumor metastasis and parasitic infection, using proteins that
PT   contain Ena/VASp binding sites.
XX
PS   Claim 13; Pages 98-100; 109pp; French.
XX
CC   The present invention relates to a method for identifying modulators of
CC   actin polymerisation. The method involves using proteins that contain at
CC   least one binding motif for proteins of the Ena/VASP
CC   (vasodilator-stimulated phosphoprotein) family in the preparation of
CC   reagents for identification/screening of molecules that modulate
CC   formation of the actin cytoskeleton. The proteins used in the method
CC   (i.e. the proteins with binding motif(s) for Ena/VASP proteins) do not
CC   bind to the Arp2/3 protein complex. The modulators identified by the
CC   method are potentially useful for treating disorders of actin
CC   polymerisation, e.g. metastatic cancer or parasitic infection; and as
CC   cytotoxic agents. The present sequence one such protein with binding
CC   motif(s) for Ena/VASp proteins, which was used in the method of the
CC   present invention.
XX
SQ   Sequence   559 AA;

US-10-001-857-42 (1-3096) x AAM52317 (1-559)

Alignment Scores:
Pred. No.:
Score:                  0.00021
Percent Similarity:     145.00
Best Local Similarity:  40.52%
Query Match:            31.37%
DB:                     2.64%
                        22

                Length:        559
                Matches:       48
                Conservative:  14
                Mismatches:    61
                Indels:        30
                Gaps:           8

Qy     559 CCATTATGCCTA------------------CTTCTCCCGTGGTCACCCGCGCCGCACGCACT 515
           |||.||||.||||                 ||:|||:||||||||||||||||||||||||
Db     294 ProThrCysIleSerSerAlaThrGlyLeuIleGluAsnArgProGlnSerProAlaThr 313

Qy     514 GTCCGGACCACGGAAATAACGACGCGCGCCTCGGCGCGCGCCCCCCTTCCCTTCTCAGCC 455
           ||||||||||||||:||||||||||||||||||||||||||||||||:||
Db     314 GlyArgThrProValPheValSerProThrProProProProProProLeuProSer 333

Qy     454 CAGCCCGACCGGCAGCGTGCACGCAATGCGTATGCCCGGACCGGACGACCCC-----GCC 401
           |||||||||||||||||||||||||:|||||||||:|||||||||||||||     |||
Db     334 AlaLeuSerThrSerSerLeuArgAlaSerMetThrSerThrProProProValPro 353

Qy     400 CCCTCTCCCTTACTCCCGGGGGACTTCG---CGCTCCCGCGGCGCCACACCCCCT------ 353
           ||||| ||||||||||:|||:|||||||   |||:||:||||||||||||:||||
Db     354 ProProProProProAlaThrAlaLeuGlnAlaProAlaValProProProProAla 373

Qy     352 -------------------------------------CCCTCGGCCACCG----------CCTCCGTTC 332
           .|:||||||||||||    |||||||
Db     374 ProLeuGlnIleAlaProGlyValLeuHisProAlaProProProIleAlaProProLeu 393
```

Search completed: November 25, 2003, 03:13:44
Job time : 122.5 secs

```
Qy     331 GGCCGGATGGGCGGGCGGGCGGCCCTTCTCTTCTCGCGAGATTAGAGACCGATTCGGCCGAGACCGGAAGTT 272
           |||:|                  |||||||||||||||||||||:|:||||||||      |||
Db     394 ValGlnProSerProPro-------ValAlaArgAlaAlaProValCys------GluThr 409

Qy     271 ATCTTTTTTCACCCTCTGTCGGAGTCCTGGTAGTGGGGCGGTTCCCGGAAAGAGGCGGAG 212
           :::  |||||||||||    :::
Db     410 ValProValHisProLeuProGlnGlyGluValGlnGlyLeuProProProProProPro 429

Qy     211 CCCGGAGTCTCAGAGCCCGCCCGCCCCTGCGCGTGCCCTCT 173
           ||||                 |||       ||||||||
Db     430 ProProLeu-----ProProGlyIleArgProSer 440
```